A GRAPH LIMIT APPROACH TO SERIATION

by

Teddy Mishura

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Mathematics

Spring 2023

A GRAPH LIMIT APPROACH TO SERIATION

by

Teddy Mishura

Approved: ____
Mark Gockenbach, Ph.D.
Chair of the Department of Mathematical Sciences

Approved: ____
John Pelesko, Ph.D.
Dean of the College of Arts and Sciences

Approved: ____
Louis F. Rossi, Ph.D.
Vice Provost for Graduate and Professional Education and
Dean of the Graduate College

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Mahya Ghandehari, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Dominique Guillot, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Sebastian Cioaba, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Jeannette Janssen, Ph.D.
Member of dissertation committee

ACKNOWLEDGEMENTS

I wish to thank my adviser Mahya Ghandehari for her support, my friends Sarah Tillman and Adam Callanan for their editing help, and my partner Charli Klein, whom I could not have done this without.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

The study of graph limit theory began in earnest in 2006 with the publication of the paper ***Limits of dense graph sequences*** by Lovász and Szegedy. Through the lens of homomorphism densities—the probability that a random map from a fixed graph F to a graph G is a homomorphism—this paper introduced new tools to the mathematical world that allowed sufficiently large networks to be viewed as random samples from a suitably chosen symmetric function $w : [0,1]^2 \to [0,1]$ called a ***graphon***. Furthermore, Lovász and Szegedy showed that this method of convergence is equivalent to convergence in a specific norm known as the ***cut norm***, where graphs are embedded into $[0,1]^2$ as step function in the form of their adjacency matrices and the distance is calculated between these step functions. This allows for graph theoretical questions to be translated into questions about step functions, where more standard tools of analysis can be employed. Furthermore, in the reverse direction, one can study how analytical properties of the function can affect the combinatorial properties of its sampled graphs.

We address this question in the context of ***geometric graphs***, whose edge structure is derived from a linear embedding in $\mathbb{R}$. This forces their adjacency matrices to increase toward the main diagonal, also known as the ***Robinson*** property, and one that easily translates to graphons. Given a convergent sequence of graphs that become increasingly close to geometric graphs—in the sense of the cut norm—can one claim the limiting object must also be Robinson? This question was solved completely in the affirmative for dense graph sequences, which are sequences of graphs G_n with positive edge density $|E(G_n)|/n^2$. We therefore focus on graph sequences that are not dense, e.g. whose edge density tends to 0, which corresponds to studying graphons that are unbounded but have finite p-norm. Such graphs are often referred to as being ***sparse***.

Specifically, we introduce and investigate a graph parameter Λ that measures by how much a graphon "fails" to be Robinson, and which is continuous with respect to the cut norm. We also develop a method that constructs Robinson approximations of L^p graphons such that the difference in cut norm between the original graphon and the approximation is dependent on Λ of the original; thus, the closer to being Robinson the original graphon is, the better the approximation.

Chapter 1

INTRODUCTION

One of the fundamental problems in modern graph theory is that of understanding the behaviour of large networks. Given a simple graph on a large number of vertices, how does one compare it to another large graph? A natural response to this can be found in [67] via the comparison of ***homomorphism densities*** $t(F, G)$, the probability that a random mapping between the vertex set of a simple graph F to the vertex set of a simple graph G is a homomorphism (i.e., preserves adjacencies). For two graphs G and H, if $t(F, G)$ is close to $t(F, H)$ for aptly chosen graphs F, then it is reasonable to say that G and H must also be close. Another natural response in the attempt to understand the behavior of large graphs is to consider a sequence of graphs whose number of vertices tend to infinity and determine if these graphs "converge" to a meaningful limit.

Lovász and Szegedy showed in [67] that for sequences of dense graphs, such limit objects do exist, in the form of symmetric and measurable functions $w : [0,1]^2 \to [0,1]$ called ***graphons***. It is sometimes useful to think of bounded symmetric measurable functions $v : [0,1]^2 \to \mathsf{R}$; we call these ***kernels*** and they can be thought of as linear combinations of graphons. Every graph G can be associated to a graphon w_G by way of its adjacency matrix, and furthermore the graphons w are the set of limit points for these graphs. Every graphon w can be used to construct a random graph model whose samples are graphs of any desired size. To form such a sample, one simply takes n independent random variables $x_1, ..., x_n$ that are uniformly distributed on $[0,1]$ and connects vertices i and j with probability $w(x_i, x_j)$.

We can think of the homomorphism density of a graph F mapped into a graphon w by considering the limit of $t(F, G_n)$ as n goes to infinity, where the G_n are graphs on

n vertices sampled from w. We furthermore think of convergence of graph sequences in the sense of homomorphism densities, in that for a graph sequence $\{G_n\}$ we say it is convergent if for all simple graphs F, the sequence of real numbers $\{t(F, G_n)\}$ is convergent itself. It was shown in [67] that if $\{G_n\}$ is convergent then there exists some graphon w such that $t(F, G_n) \to t(F, w)$ for all F; furthermore, every graphon w' has some sequence of graphs $\{G'_n\}$ such that for all F, $t(F, G'_n) \to t(F, w')$.

Thus, if one can show that a graph and a graphon are close in some sense, then it must be that samples of the graphon should be close to the graph in question. However, it is impractical to calculate every homomorphism density for a given graph; luckily, there exists an analytical workaround. The ***cut norm***, introduced in [38] and defined by

$$\|w\| \; = \sup_{A,B \subseteq [0,1]} \left| \iint_{A \times B} w(x, y) dx dy \right|,$$

has a close relation with the theory of dense graph limits. If $\|w - u\|$ is small, then $t(F, w)$ must be close to $t(F, u)$ for all F, and vice versa. Convergence in the cut norm is therefore equivalent to convergence of every sequence of homomorphism densities. When thinking of the distance between graphs, we simply use their associated graphons (e.g. $\|G - H\| \; := \|w_G - w_H\|$). This also allows for a natural way to think of the distance between a graph G and a graphon w. The cut norm also respects sampling from random graph models; for a growing sequence of graphs $\{G_n\}$ sampled from a graphon w, almost surely, there exists a vertex labelling ϕ_n of the graphs $\{G_n\}$ such that $\|w_{G_n^{\phi_n}} - w\| \; \to 0$ [12, Proposition 5.2]. The cut norm is therefore viewed as the correct metric with which to define graph convergence. Importantly, this implies that graphons form the background reality of finite simple graphs, as every graph can be thought of as a sample from a parent graphon, whose properties can be studied with the familiar tools of analysis. Dense graph theory is discussed more in-depth in Section 2.2.

Dense graph theory mandates that any sparse graph sequence $\{G_n\}$ (i.e. $|E(G_n)|/n^2 \to 0$) must converge to the identically zero graphon. This is because the homomorphism density of F in G_n tends to 0 for any simple graph F with at least one edge if $\{G_n\}$ is a sparse sequence. To combat this uninteresting result, multiple versions of sparse graph convergence have been developed by different authors. The first sparse theory to arise was Benjamini-Schramm convergence (see [11, 45]) to study convergence of sequences of graphs with uniformly bounded maximum degree (also called very sparse). Limit objects of such sequences are called ***graphings*** and their study was motivated by the random graph models developed in [10].

It is unfortunately the case that uniformly bounded maximum degree is a restrictive condition that excludes many graph sequences that are not dense. In [12, 13], Borgs et al. lay the framework for a notion of convergence that provides nontrivial results for sparse graph sequences. This is the primary method of sparse convergence studied in this thesis; it is done by normalizing the graphs G_n by their edge density and considering their convergence in the cut distance. Such normalization of a sparse graph sequence will force the associated limit graphon to become unbounded, thus, the limit object will be an element of $L^p([0,1]^2)$ for some $1 \leq p < \infty$ rather than $L^\infty([0,1]^2)$. We refer to these limit objects as L^p ***graphons***: symmetric measurable real-valued functions in $L^p([0,1]^2)$. We note that normalizing graph sequences by their p-norm is too strong a condition for $p > 1$; this is because, if $\{G_n\}$ is a sequence of sparse graphs, dividing w_{G_n} by $\|w_{G_n}\|_p$ for $p > 1$ will not change that the edge density of the sequence will still tend to zero, resulting in a trivial limit.

Many results from dense theory generalize to sparse theory. Importantly, every L^p graphon w gives rise to a random graph model whose samples are (sparse) graphs of any desired size. Similar methods used in dense theory can be combined with these models to show that sparse graph sequences can be viewed as samples from an L^p graphon, rather than the trivial graphon. Sparse (and very sparse) graph limit theory is discussed more thoroughly in Section 2.3.

Graph limit theory is a wide field with many notions of convergence; while the results of this thesis only utilize dense and sparse graph limit theory, recently there have been several advances in the field that extend and generalize these convergence models.

In [55], the authors introduce a metric that allows the limit objects of both dense and sparse graph sequences under this metric to both be represented by symmetric Borel measures on $[0,1]^2$. Additionally, such sequences of graphs are special examples of limit objects in this new framework, showing its rich potential for advancement of the field. We delve into extensions of dense and sparse graph limit theory in more detail in Section 2.4.

This thesis focuses on the intersection between sparse graph limit theory and ***geometric graphs***—these are graphs that exhibit a spatial, line-embedded structure. See [15, 47] for examples of such latent space models for social networks. In these models, vertices can be identified with points on the line segment $[0,1]$, and the edges are created between points $x < y$ with increasingly large probability as y moves closer to x. Specifically, for the adjacency matrix A_G of such a graph, for all $i < j < k$,

$$(A_G)_{i,k} \leq \min\{(A_G)_{i,j}, (A_G)_{j,k}\}. \tag{1.1}$$

A symmetric matrix A with the property displayed in (1.1) is called a ***Robinson*** matrix; it is called ***Robinsonian*** if it becomes a Robinson matrix after simultaneous application of a permutation π to its rows and columns. Such a permutation induces a ***Robinson ordering*** of A.

The Robinson property translates quite naturally to graphons; an L^p-graphon $w : [0,1]^2 \to \mathsf{R}$ is called a ***Robinson*** graphon if for all $x < y < z$,

$$w(x,z) \leq \min\{w(x,y), w(y,z)\}. \tag{1.2}$$

Likewise, an L^p-graphon w is called ***Robinsonian*** if there exists a measure preserving bijection $\phi : [0,1] \to [0,1]$ such that $w^\phi = w(\phi(x), \phi(y))$ is Robinson. In this more general case, the bijection ϕ is the natural extension of the permutation π used for Robinsonian matrices. Robinson graphons were introduced in [19] under the name of diagonally increasing graphons. These are generalizations of the Robinson matrices

$$
\begin{bmatrix}
0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0
\end{bmatrix}
\qquad
\begin{bmatrix}
0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0
\end{bmatrix}
$$

Figure 1.1: An example of a Robinsonian 10x10 matrix and its Robinson ordering. The left version is improperly ordered, while the right version has been properly ordered.

(also known as R-matrices) that appear in the study of the difficult problem of ***seriation***, which deals with reordering objects into a sequence that best reveals regularity and patterning among the whole.

Seriation was first written about in a systematic fashion by Petrie in [78], but its evaluation remained primarily intuition-based until Robinson [81] and Brainerd [14]. It was further popularized in archaeology by Kendall [53]. This formalization paved the way for seriation to be viewed as a mathematical problem whose objective is to order a set of items so that given some similarity ranking, similar items are placed closer to each other than dissimilar items. This translates naturally into a question regarding symmetric matrices in the following way: If the entries $A_{i,j}$ of A represent similarity between i and j, then a Robinson ordering represents a linear arrangement of the items so that similar items are placed close together (see Figure 1.1).

Since its introduction in archaeology, seriation has also been studied and developed in many fields. In cartography and graphics, emphasis was put on rigorous data visualization in regards to seriation by Bertin [9], leaving a lasting impact on the discipline. This can be seen in the works of Mueller et al. [75] and Ling [64] where the authors proposed new methods for visualizing similarity matrices and large scale graphs, as well evaluating the interpretability of results given by different vertex ordering algorithms. Other fields also have their own discussions of how to visualize and quantify similar data, including sociology and sociometry [34, 25], psychology and psychometrics [48, 49], ecology [72, 54], biology and bioinformatics [84, 44, 16], and operations research [71, 61], to name a few. We refer to [62] for a more in-depth analysis of the historical perspective of seriation.

Recognizing when a matrix is Robinsonian—as well as finding a proper Robinson ordering—is a problem that can be solved in polynomial time. The original polynomial time algorithm for this seriation problem can be found in [73], while several more modern and efficient algorithms can be found in [4, 37, 58, 56]. There has also been recent interest in a generalization of the ***linear seriation*** problem known as ***circular seriation***, where the goal is to find a circular order of the objects of interest in a manner consistent with their pairwise similarities. An optimal polynomial time algorithm for determining such an ordering is given in [3]. Many of these algorithms rely upon graph theoretical knowledge, where they relate Robinson matrices to the adjacency matrices of unit interval graphs, or on spectral methods, where a specific eigenvector must be computed and properly ordered. The spectral method, in particular, was used in [32] applied to the ***ranking problem***, a similar but distinct problem from seriation. In [42], an algorithm using cutting planes was developed for seriation of certain specific classes of input-output matrices. However, if a matrix is "almost" Robinsonian—perhaps a Robinsonian matrix perturbed by noise—the results of these algorithms are at worst meaningless and at best poorly understood.

In this situation, the best solution is to find an "almost" Robinson ordering of the original matrix: an ordering that minimizes the distance between the matrix and the set of Robinson matrices in some norm. This is a challenging problem and was shown in [18] to be NP-hard for finding such an ordering that minimizes the ℓ^∞ distance; a factor 16 approximation for this problem was given in [18]. A number of problems related to Robinson ordering approximation in ℓ^p distance for integer p were shown to be NP-hard in [6]; specifically, it was shown that finding a ***proper strong Robinson relation***, a stronger condition than being Robinsonian, in the ℓ^p distance is NP-complete. A statistical approach to the noisy seriation problem is developed in [31] where distance in the Frobenius norm is considered; however, their work is interested in finding orderings of matrices that approximate column properties such as monotonicity or unimodality, rather than two dimensional properties such as being Robinsonian. The work in [33, 77, 70] studies recovering almost Robinson orderings for noisy matrices with respect to the Frobenius norm. For work involved recovering almost Robinson orderings for noisy matrices with less standard norms, we invite the reader to peruse [69] for approximations using the Kendall tau norm.

When the ordering of the matrix's rows and columns are fixed, the problem of Robinson approximation becomes that of finding a Robinson matrix who is closest to the given matrix in some chosen norm. This problem is well known to be solvable in polynomial time for the ℓ^∞ norm, as the explicit closed form of the optimal solution can be obtained [82]. While not optimal, an ℓ^1 approximation that can be computed in polynomial time was found in [40]; moreover, for binary matrices the approximation can be itself binary and the results will still hold. In terms of the cut norm, an approximation was found in [19], but it cannot be computed in polynomial time. However, it is the first cut norm approximation for the Robinson problem that this author knows of; we expand upon their results in this thesis. Seriation and its history are talked about more at length in Section 2.1.

Most graphs that appear in practical applications are large; unfortunately, large graphs are not likely to be geometric graphs. Equivalently, matrices of large sizes will likely fail to be Robinson, as the condition imposed by Equation (1.1) is too restrictive. However, one expects random instances of Robinson graphons to be "almost Robinson" if the sample size is large. That is, in the sampled graph most edges are expected to occur between vertices whose labels are close. Note that if a Robinson graphon w is not $\{0, 1\}$-valued, then it may produce, with high probability, graphs that are not geometric graphs. So, in order to recognise samples of Robinson graphons, we need to utilize a graph/graphon parameter (applied to labelled graphs) which acts as a gauge of Robinson property. Importantly, this parameter should also be continuous in the cut norm, as to reliably measure behavior of graph samples, it must respect graph limits.

One's first thought for this problem is to simply compare every entry in the matrix with its preceding neighbors and sum the differences of the entry further from the diagonal is greater than the entry closer to the diagonal. Clearly, a matrix will only be Robinson if this process results in 0. In [39], the authors defined a parameter Γ_1 that does this, and while Γ_1 is simple to compute, it fails to be continuous in cut norm (or, equivalently, the graph limit topology). Thus Γ_1 is not a suitable Robinson measurement for growing networks, as it does not respect limits of graph sequences. However, Γ_1 is a suitable measurement for the Robinson property of a single matrix of a fixed size, as it can be computed in polynomial time. Additionally, it is resistant to noise, as [39] shows that matrices with a small Γ_1 value are close to a Robinson matrix in the normalized ℓ^1 norm. Γ_1 therefore fills a certain niche, being useful for the study of a matrix of a fixed size rather than a growing sequence.

For a cut norm continuous parameter that characterizes the Robinson property, we look to [19, 40], where they introduced a function Γ on the space of L^∞ graphons. Γ was shown to be a suitable measurement for the Robinson property; additionally, the authors used Γ to formalize the idea of ***almost Robinson graphs***, which are large graphs sampled from Robinson graphons. This is done through the use of the following two properties: firstly, $\Gamma(w) = 0$ precisely when w is a Robinson graphon [19], and secondly, w is close to a Robinson graphon if and only if $\Gamma(w)$ is close to 0 [19, 40], where the idea of closeness between graphons is measured by the difference in cut norm. This allows Γ to act as a continuous measure of the Robinson property whose output has quantitative value. Both Γ and Γ_1 are covered in more detail in Section 4.1.

In this thesis, we aim to characterize almost Robinson graphs for the case of sparse graph sequences; that is, we look for L^p graphon parameters which would satisfy both properties listed above. We first show that the sparse problem cannot be resolved just as an automatic extension of the previous work on Γ, initiated in [19]. Namely, we show that the argument in [19] used to prove the $\| \cdot \|$ -continuity of Γ cannot be extended to sparse graphs/unbounded graphons because their proof relies upon the fact that the triangular cut operator—multiplying a graphon by 1 above the main diagonal and by 0 below the main diagonal—is bounded in cut norm for graphons in

$\mathcal{W}^\infty$. They were unable to show whether this method of proof could be extended to the sparse case, as it was unknown whether the triangular cut was bounded in cut norm for graphons in $\mathcal{W}^p$ for $p < \infty$. To answer this question, we can look toward matrices for inspiration, seeing how a similar triangular cut would behave for them in cut norm. Because symmetric matrices can be thought of as a dense subspace of graphons, we can prove results in the world of matrices to then make statements in the world of graphons.

In Chapter 3, we study the norm of the triangular cut on matrices with respect to the cut norm. It is not unusual to see the cut norm appear in a study of operators; while the cut norm arose from the graph theoretical problem of finding the maximal cut of a given weighted graph, it has been of much use in more analytical pursuits, such as approximation results for graphons [68] and algorithmic approximations for matrix problems [1, 38]. Consider an $n \times n$ matrix and let the operator $\mathcal{T}_n$ set all entries of that matrix below the main diagonal to 0. $\mathcal{T}_n$ is called the triangular cut (also known as triangular truncation), and its norm growth is a well-studied problem in operator theory; see for example [2, 89] for explicit calculations and bounds for $\mathcal{T}_n$ applied to matrices equipped with operator norm.

Outside of the cut norm, it is well known that the operator norm of $\mathcal{T}_n$ grows to infinity when it is considered as an operator on real matrices equipped with the standard operator norm $\|\cdot\|_{\mathrm{opr}}$ (see [2, Theorem 1] for a proof and [88] for estimations of growth speed). In [8], Bennett showed that when $1 < p < q < \infty$, the triangular cut mapping ℓ^p to ℓ^q is bounded. Recently, Coine used the canonical characterization of Schur multipliers—an operator that multiplies one matrix element-wise with another matrix—to prove that the triangular cut mapping ℓ^p to ℓ^q is unbounded when $1 \leq q \leq p \leq \infty$ [20].

However, to make conclusive statements about graphons we must know how the norm of $\mathcal{T}_n$ grows on symmetric matrices. The author could not find any such analysis in the literature, with respect to the cut norm or otherwise, and so we address this issue here. To do so, we view $\mathcal{T}_n$ as a Schur multiplier and further note that the cut norm is equivalent to an injective tensor norm. This allows us to make use of some bounds and techniques from [7] to show that the norm of $\mathcal{T}_n$ on symmetric matrices grows to infinity. We refer to [7, 20, 23] for results on the norms of Schur multipliers in general. Due to the author's interest, we also present a proof of the unboundedness of $\mathcal{T}_n$ on symmetric matrices with respect to the standard operator norm.

As mentioned previously, the parameters Γ_1 and Γ are useful measurements of the Robinson property in specific circumstances. However, due to the reasons given above, their definitions were not extended to sequences of sparse graphs, which correspond to sequences of L^p graphons. A new parameter was therefore needed to measure the Robinson property of such sequences of graphs/graphons. Utilizing new ideas about how to measure Robinson properties of a function in $[0, 1]^2$ alongside a key lower bound for the function Γ [40, Lemma 4.1], the parameter Λ was developed. This new parameter is an accurate and useful measure for the Robinson property for ***all*** sequences of convergent graphs/graphons, sparse or dense.

The generating idea behind the creation of Λ was that the Robinson property also affects the local average values of a graphon alongside the pointwise values. If a graphon is Robinson, then the average value of the graphon near the diagonal should be larger than the average value of the graphon further from the diagonal. In this way, one can measure the Robinson property by comparing "blocks" of the graphon, rather than a comparison of each value the graphon achieves. This way of viewing the Robinson property through integration rather than pointwise comparison is the key factor of why Λ can be so easily extended to L^p graphons; it also allowed us to make use of several estimation and averaging techniques to prove strong results about Λ.

In Chapter 4.2, we state and prove various results about Λ. We begin by showing that it identifies the Robinson property and that it is continuous with respect to the cut norm, for L^p and traditional graphons alike. However, more importantly, we show that Λ is ***Robinson stable***—i.e., it can be used to create a "good" Robinson approximation of a non-Robinson graphon. Here the quality of an approximation is measured by its distance from the original in cut norm. This achieves the important goal of showing that Λ truly does measure "how Robinson" a graphon is, for if $\Lambda(w)$ is close to 0, you can identify a Robinson graphon u such that $\|u - w\|$ is small.

This result was significantly more difficult to show; while the framework of the proof of stability of Γ in [40] was used, several new ideas were required to extend this proof to accommodate L^p graphons. Additionally, Λ was shown to have more accurate Robinson approximations than Γ for L^p graphons with $p > 5$. In a sense, this implies that Λ more accurately measures when a graphon is "close" to being Robinson—if $\Lambda(w)$ is small, it is more meaningful than $\Gamma(w)$ being small. Λ is thus a true successor to Γ, providing closer Robinson approximations for both traditional and L^p graphons. It also has a far simpler definition that lends itself to easier computation.

Chapter 2

BACKGROUND

2.1 Seriation

Since it was mathematically formalized by Robinson in [81], seriation—the study of optimally ordering objects according to similarity or ranking—has enjoyed a rich history of study in a variety of fields. We shall focus on primarily mathematical research attempting to solve the linear ordering problem. In this section, we will introduce some of the results currently present in the field as motivation for our own research, beginning with one of the more well-studied topics in seriation: recognition algorithms for Robinsonian matrices.

2.1.1 Recognition algorithms

In this subsection, we outline two types of algorithms used in the seriation problem; firstly, methods for determining if a given matrix is Robinsonian, and secondly, methods for determining a Robinson ordering for a given Robinsonian matrix. We begin by recalling the definition of a Robinsonian matrix.

Definition 2.1.1 (Robinsonian matrix)**.** A symmetric matrix A is called ***Robinsonian*** if it becomes a Robinson matrix after simultaneous application of a permutation π to its rows and columns. Such a permutation induces a ***Robinson ordering*** of A.

While most methods the author is aware of are primarily concerned with the latter of the two types of algorithm, there has been recent interest in finding characterizations of Robinsonian matrices to determine if the seriation problem is well posed. In [58], Laurent et al. develop a method to certify Robinsonian matrices by using graph theoretic knowledge, as the adjacency matrix of a graph is Robinsonian exactly when a graph is a unit interval graph. As a reminder, a unit interval graph is an undirected graph formed from a set of unit intervals on the real line, with a vertex for each interval and an edge between vertices whose intervals intersect. The authors then generalize the well-known graph theoretic characterization of unit interval graphs via forbidden substructures to matrix theory; in particular, they extend the notion of an ***asteroidal triple*** to matrices and use this to characterize Robinsonian matrices.

We recall some definitions; if G is a graph, then for $x, y, z \in V(G)$, a ***path from x to y missing z*** is a path $P = (x = x_0, x_1, \ldots, x_k, x_{k+1} = y)$ in $V(G)$ such that all pairs $\{x_i, x_{i+1}\}$ are adjacent to each other and x_i is not adjacent to z for all $0 \leq i \leq k$. An ***asteroidal triple*** is a set of vertices $\{x, y, z\} \subseteq V(G)$ such that they are pairwise not adjacent and between every two of the vertices there exists a path connecting them that misses the other. These were introduced in [60] as forbidden substructure characterizations of interval graphs, where it was shown that a graph is an interval graph if and only if it is chordal and if it contains no asteroidal triples. As a graph is a unit interval graph if and only if it is a claw-free interval graph (see [43, pp. 139-146]), the following characterization must hold.

Proposition 2.1.2. *A graph G is a unit interval graph if and only if it satis es the following conditions:*

(i) *G is chordal; speci cally, G does not contain an induced cycle of length at least 4.*

(ii) *G is claw-free; speci cally, G does not contain an induced claw $K_{1,3}$.*

(iii) *G contains no asteroidal triples.*

Laurent et al. then proceed to generalize the idea of an asteroidal triple to the more general setting of matrices to obtain a similar forbidden structural characterization of Robinsonian matrices as Proposition 2.1.2. They begin with the notion of a path that misses an entry in a matrix. If A is a matrix indexed by some finite set V, then given $z \in V$, a ***path avoiding z in A*** is of the form $P = (v_0, \ldots, v_k) \in V^{k+1}$ where all triples (v_{i-1}, z, v_i) are such that $A_{v_{i-1},v_i} > \min(A_{v_{i-1}z}, A_{v_i,z})$. This is referred to in

[58] as being ***not Robinson*** and is used to extend the definition of an asteroidal triple as shown below.

Definition 2.1.3 (Weighted asteroidal triple)**.** Let A be a symmetric matrix. A triple $\{x, y, z\}$ of distinct elements of V is called a ***weighted asteroidal triple*** of A if there is a path between each pair of elements that avoids the third.

The main result of [58] is that weighted asteroidal triples are the only substructures that need to be forbidden to achieve the Robinsonian property.

Theorem 2.1.4 (Theorem 1.3, [58])**.** *A symmetric matrix A is Robinsonian if and only if it does not contain any weighted asteroidal triples.*

The authors construct an algorithm to enumerate all weighted asteroidal triples present in a matrix; if this returns 0, the matrix is therefore Robinsonian. We describe the algorithm as follows. Given a matrix A indexed by a finite set V and distinct entries $x, y, z \in V$, it is possible to check whether there exists a path from x to y avoiding z in $O(n^2)$ time. This can be done by considering the graph H_z with vertex set $V \setminus \{z\}$ where two vertices $u, w \in V \setminus \{z\}$ are adjacent if and only if $A_{uw} > \min(A_{uz}, A_{wz})$. Building the graph H_z and then checking if there exists a path between x and y in H_z can be done in $O(n^2)$ time. An elementary approach would be to test all possible triples using this method for an algorithm that would run in $O(n^5)$ time. However, the following simpler algorithm introduced in [58] runs more efficiently, in $O(n^3)$ time.

Given a symmetric matrix A indexed by a finite set V, initialize a function $f : \binom{V}{3} \to \mathsf{Z}$ by $f(\{x, y, z\}) = 0$ for all $\{x, y, z\} \in \binom{V}{3}$. Then, for each $v \in V$, construct the graph H_v as described above. If $\{x, y\}$ is a pair of elements lying in the same connected component of H_v, set $f(\{x, y, v\}) \to f(\{x, y, v\}) + 1$. If at the end of this process there exists a triple $\{x, y, z\} \in \binom{V}{3}$ such that $f(\{x, y, z\}) = 3$, then this forms a weighted asteroidal triple. Otherwise, A has no weighted asteroidal triples and is thus Robinsonian. This process completes in $O(n^3)$ time.

More recently, similar research by the same authors has been done in [59] to characterize matrices of chordal graphs, i.e. graphs with no induced cycles of length at least 4. A graph is chordal if and only if it has a ***perfect elimination ordering***; the authors generalize this concept to matrix theory in the form of ***weighted chordless walks***. These are then shown to be the determining substructure for chordal graphs; that is, a symmetric matrix A has a perfect elimination ordering if and only if it has no self-contained pair of weighted chordless walks. We now turn our attention to algorithms that are designed to find a proper Robinson ordering given a matrix already known to be Robinsonian. While this problem can be tackled from many different avenues, the relationship between Robinson matrices and interval graphs has proven especially fruitful for this endeavor, with many algorithms based on graph theoretical/combinatorial results being developed in recent years (see [57, 35, 56, 3, 51] for examples). We shall discuss the results shown in [35] and [56] specifically, as both have been the subject of recent improvement ([36] and [57, 59] respectively) and give important background to the techniques commonly used in the field. We note that as an algorithm that achieves this would need to check each entry in the matrix at least once, no algorithm can run in time less than $O(n^2)$ for a given $n \times n$ matrix. Both mentioned algorithms achieve the optimal time of $O(n^2)$, as we shall show below, starting with the algorithm defined in [35]. As it relies upon clever navigation of specifically constructed PQ-trees, we begin with their definition.

Definition 2.1.5 (PQ-tree)**.** A PQ ***tree*** $\mathcal{T}$ on a set S is a tree that represents a set of permutations on S. A PQ tree represents its permutations via permissible reorderings of the children of its nodes. The children of a P node may be reordered in any way, while the children of a Q node may be put in reverse order, but may not otherwise be reordered. A PQ tree represents all leaf node orderings that can be achieved by any sequence of these two operations. For a vertex α of a PQ tree $\mathcal{T}$, we denote by $\mathcal{T}(\alpha)$ the subtree of $\mathcal{T}$ with root α; by S_α, we denote the set of leaves of $\mathcal{T}_\alpha$. For an example of a PQ tree, see Figure 2.1

Given a symmetric matrix A indexed by some finite set S, for $x \in S$, define the

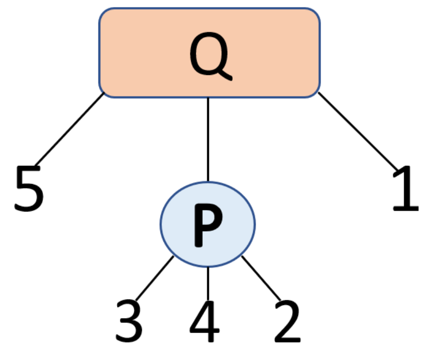

Figure 2.1: A PQ tree representing permissible permutations of the set $\{1, 2, 3, 4, 5\}$.

sets $\Gamma_k(x)$ for $k \in \{1, \ldots, n\}$ as follows:

$$\Gamma_0(x) = \{x\}$$
$$\Gamma_k(x) = \{y \in S \mid \forall z \in S \setminus \Gamma_{k-1}(x),\ A_{x,y} \leq A_{x,z}\}$$

$\Gamma_k(x)$ are "balls" centered at x introduced for the following reason.

Proposition 2.1.6 (Claim 2, [35])**.** *For a symmetric matrix A indexed by some finite set S, if A is Robinsonian and S is sorted into a Robinson ordering, then the sets $\Gamma_k(x)$ are intervals.*

We now additionally define the following family of graphs G_k for $k \in \{1, \ldots, n\}$ and a given symmetric matrix A:

Definition 2.1.7 (Intersection graphs)**.** For every $k \in \{1, \ldots, n\}$, let the intersection graph $G_k = (V_k, E_k)$ be defined as

$$V_k = \{\Gamma_j(x) \mid x \in S,\ 0 \leq j \leq k\}$$
$$E_k = \{\{v_i, v_j\} \mid v_i \cap v_j \neq \emptyset\}.$$

When A is Robinsonian, the graphs G_k are interval graphs. We also note that the PQ-tree of G_n represents the set of permutations compatible with A alongside the following fact: If α is a vertex of the PQ-tree of G_n such that $x, y \in S_\alpha$ and $z \in S \setminus S_\alpha$, then $A_{x,y} \leq A_{x,z} = A_{y,z}$. Combining these facts allows for the basic skeleton of the algorithm to be built, which we will outline below.

- Pick some fixed number K that is independent of S (what value this is has no bearing on the algorithm). Then, check if the graph G_K is an interval graph. If it is not, stop the process. If it is, build its PQ-tree which we shall call $\mathcal{T}_K$.

- For each P-vertex α in $\mathcal{T}_K$, check if $A_{x,y}$ is the same value for $x \in S \setminus S_\alpha$ and $y \in S_\alpha$. If this is not so, transform x into a Q-vertex.

- For each Q-vertex α, check if $x, y \in S_\alpha$ and $z \in S \setminus S_\alpha$ implies that $A_{x,y} \leq A_{x,z} = A_{y,z}$. If this is not the case and α is the child of a Q-vertex, delete it, with its children becoming children of its parent.

  After this step, if A is Robinsonian, we should have subsets of S which are intervals for any compatible permutation. These intervals will correspond to vertices of $\mathcal{T}$.

- For all P-vertices α such that $|S_\alpha| > K$, construct a set $S'_\alpha \subset S_\alpha$ such that S'_α contains the following elements:

  (i) The first and last element of each Q-vertex β such that $|S_\beta| > K$ and β is a child of α.

  (ii) One element (it does not matter which) of S_γ for each P-vertex γ such that $|S_\gamma| > K$ and γ is a child of α.

  Recursively determine if A restricted to S'_α is Robinsonian and construct its PQ-tree $\mathcal{T}'$. Stop the recursion when $|S'_\alpha| < 3$. Merge α and $\mathcal{T}'$.

- Choose an arbitrary permutation given by $\mathcal{T}$ and verify if it corresponds to a Robinson ordering. If so, A is Robinsonian; otherwise, it is not.

During the first step, the algorithm considers some values of A (those that determine G_K) and constructs, if G_K is an interval graph, the PQ-tree $\mathcal{T}$ which represents all the permutations compatible with these values. Because $\mathcal{T}$ corresponds to a subset of entries of A, it represents a set of permutations which contains all the permutations compatible with A. The second through fourth steps consider more and more values of A and update $\mathcal{T}$ such that $\mathcal{T}$ is always representative of the set of the permutations compatible with the considered values of A. At every moment, $\mathcal{T}$ represents a superset of the permutations compatible with A. Thus, at the end of the fourth step, if A is Robinsonian, the entries of A not used in the algorithm have no effect on $\mathcal{T}$, and thus $\mathcal{T}$ represents the set of the permutations compatible with A. Any permutation chosen from $\mathcal{T}$ should then correspond to a proper Robinson ordering for A; if it does not, then A is not Robinsonian.

This process terminates in $O(n^2)$ time in both worst and best case, as it must check the re-ordered matrix to see if it is truly Robinson after the chosen permutation has been applied. The developers of the previously mentioned algorithm used the well known properties of interval graphs and their relation to Robinson matrices to derive proper Robinson orderings; we discuss now another algorithm based on the properties of the more closely related ***unit interval graphs***. The performance of this algorithm scales off the number of distinct nonzero entries in a matrix and thus typically performs much better in practice. This algorithm was originally introduced in [56] and we recount its formulation here, beginning with preliminary definitions.

Definition 2.1.8 (Weak linear order)**.** An ordered partition $(B_1, \ldots, B_k)$ of a finite set V corresponds to a ***weak linear order*** ϕ on V by setting $x =_\phi y$ if x and y belong to the same block B_i, and $x <_\phi y$ if $x \in B_i$ and $y \in B_j$ with $i < j$. When all blocks B_i are singletons then ϕ is a linear order of V.

We next define the concept of a straight enumeration, as the relation between this idea and weak linear orders form the crux of this algorithm. If $G = (V, E)$ is a graph, for each $x \in V$, we let $N(x) = \{y \in V \mid \{x, y\} \in E\}$ be the neighborhood of x and $N[x] = \{x\} \cup N(x)$ be the closed neighborhood of x. Two vertices x and y are called ***indistinguishable*** if $N[x] = N[y]$; this defines an equivalence relation on V whose classes we call the ***blocks*** of G. Each block is a clique in G and two blocks B and B' are said to be ***adjacent*** if $B \cup B'$ is itself a clique in G. A ***straight enumeration*** of G is a linear ordering $\phi = \{B_1, \ldots, B_k\}$ of the blocks of G such that each block B_i is listed consecutively in the linear order with the block(s) B_j it is adjacent to. We now state a theorem of [56] relating these two concepts.

Theorem 2.1.9 (Theorem 5, [56]). *Let $G = (V, E)$ be a graph. A linear order π of V is a Robinson ordering of A_G if and only if there exists a straight enumeration of G whose corresponding weak linear order ϕ is compatible with π; i.e., satis es:*

$$\forall x, y \in V \text{ with } x \neq_\phi y \text{ we have } x <_\pi y \Leftrightarrow x <_\phi y.$$

It is thus the case that finding all the permutations corresponding to Robinson orderings of a symmetric binary matrix A can be rephrased as finding all the possible straight enumerations of the related graph G. This is achievable in linear time (see [22, 24]) and is cohesive with the fact that a similar problem known as the consecutive ones problem—finding a permutation of a binary matrice's rows such that all 1s in the columns are consecutive—can also be solved in linear time, as is shown e.g. in [27]. To extend this argument to nonbinary matrices is more difficult; we start by defining the ***level graphs***, the analogues for similarity matrices of threshold graphs for dissimilarities. If $\alpha_0 < \alpha_1 < \ldots < \alpha_L$ are the distinct values of the entries of A, then the graph $G^{(k)} = (V, E_l)$ whose edges are the pairs $\{x, y\}$ such that $A_{x,y} \geq \alpha_k$ is called the ***k-th level graph*** of A. These level graphs can be used to decompose A into a conic combination of binary matrices; moreover, A is Robinson if and only these binary matrices are Robinson [43].

If A is Robinsonian, then so must the adjacency matrices of its level graphs be. The converse does not hold, as one must find a permutation that is a Robinson ordering for the extended adjacency matrices of all the level graphs simultaneously. Roberts [43] introduced a characterization of Robinson matrices in terms of unit interval graphs that can be rephrased in the following way: A is Robinsonian if and only if its level graphs have vertex linear orders that are compatible. The authors in [56] then combine his results alongside the previously stated links between Robinsonian matrices, unit interval graphs, and straight enumerations to achieve the following characterization of Robinsonian matrices.

Proposition 2.1.10 (Theorem 7, [56]). *Let A be a symmetric $n \times n$ matrix with level graphs $G^{(1)}, \dots, G^{(L)}$. Then:*

(i) A is a Robinsonian matrix if and only if there exist straight enumerations of $G^{(1)}, \dots, G^{(L)}$ whose corresponding weak linear orders $\phi_1, \dots, \phi_L$ are pairwise compatible.

(ii) A linear order π of V is a Robinson ordering for A if and only if there exist pairwise compatible straight enumerations of $G^{(1)}, \dots, G^{(L)}$ whose corresponding common re nement is compatible with π.

Proposition 2.1.10 is the backbone of the algorithm ***Robinson***(A, ϕ) developed in [56], which we shall summarize here. First, we are given as input a symmetric nonnegative $n \times n$ matrix A and a weak linear order ϕ of $\{1, \dots, n\}$. Let G be the support graph of A; that is, the graph $G = (V, E)$ such that if $A_{i,j} > 0$ we have that $\{i, j\} \in E$.

(i) Find the connected components of G and order them in a compatible way with ϕ. If this is not possible, we stop as A therefore does not have any straight enumerations compatible with ϕ. If this is possible, initialize the weak linear order ψ.

(ii) Divide G into its connected components V_α; for each component V_α, compute the straight enumeration ϕ_α of $G[V_\alpha]$ if it exists. If it does not, we stop.

(iii) Compute the common refinement of ϕ_α with $\phi[V_\alpha]$; if this is nonempty, we continue. If it is empty, compute the common refinement of the reversal of ϕ_α with $\phi[V_\alpha]$; if this too is empty, we stop. In both cases, we denote a nonempty common refinement as ψ_α.

(iv) Take the matrix $A[V_\alpha]$; set its smallest nonzero entries to 0, obtaining a new matrix $A'[V_\alpha]$ whose nonzero entries take on fewer distinct values. If $A'[V_\alpha]$ is diagonal, then we concatenate ψ_α after $\psi_{\alpha-1}$ in the weak linear order ψ. Otherwise, make a recursive call, where the input of the recursive routine is the matrix $A'[V_\alpha]$ and ψ_α.

(v) If this successfully terminates, then the concatenation $(\phi_1, \dots, \phi_C)$ represents a straight enumeration of G and $\psi = (\psi_1, \dots, \psi_C)$ represents the common refinement of this straight enumeration with the given weak linear order ϕ and with the level graphs of A.

It is shown in [56, Lemma 12] that this algorithm runs in $O(|V|+|E|)$ time. Therefore, for sparsely populated $n \times n$ matrices it is even more efficient than algorithms that run in $O(n^2)$ time, as it saves processing power by dealing only with distinct entry values.

The prevalence of graph theoretical and combinatorial methods notwithstanding, the problem of seriation has been handled with tools from other fields as well. We shall present one of the most well known, the spectral algorithm introduced in [4], as it has been extended [30, 31] and adapted [80, 33] for use in this problem since its introduction. We begin with the fundamental definition upon which this algorithm rests.

Definition 2.1.11 (Fiedler value)**.** The ***Fiedler value*** of a symmetric, nonnegative matrix A is the smallest nonzero eigenvalue of its Laplacian $L_A := \mathrm{diag}(AJ) - A$, where J is the square matrix of entry values all equal to 1. The corresponding eigenvector is called the ***Fiedler vector*** and is the optimal solution to

$$\begin{aligned} &\text{minimize} && x^T L_A x \\ &\text{subject to} && x^T e = 0 \text{ and } \|x\|_2 = 1 \end{aligned}$$

in the variable $x \in \mathsf{R}^n$, where $e \in \mathsf{R}^n$ is the vector of entries identically equal to 1.

We now recall the main result from [4] about how the Fiedler vector relates to the Robinson ordering of a noise free Robinsonian matrix. For the sake of completeness, we include their proofs here.

Lemma 2.1.12 (Theorem 3.2, [4])**.** *If A is a Robinson matrix, then it has a monotone Fiedler vector.*

Proof. This proof uses the Perron-Frobenius theorem for its main result. We thus construct a nonnegative matrix to use this theorem on. Define the matrix $S \in \mathsf{R}^{(n-1)\times n}$

as

$$S = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}.$$

For any vector $x \in \mathsf{R}^n$, it must be that $Sx = (x_2 - x_1, \ldots, x_n - x_{n-1})^T$. We now define a matrix $T \in \mathsf{R}^{n\times(n-1)}$ by

$$T = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 1 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$

It is not difficult to show that $ST = I_{n-1}$ and that $TS = I_n - ee_1^T$, where e_1^T is the vector $(1, 0, \ldots, 0)$. We define now the matrix $M_A = SL_AT = \{m_{i,j}\}$ and let $L_A = \{l_{i,j}\}$. We now show that Sx is an eigenvector of M_A if and only if x is an eigenvector of L_A and $x \perp e$. Indeed,

$$\begin{aligned} L_Ax = & \quad \lambda x \quad \Longleftrightarrow \\ SL_Ax = & \quad \lambda Sx \quad \Longleftrightarrow \\ SL_A(I - ee_1^T)x = & \quad x \quad \Longleftrightarrow \\ SL_ATSx = & \quad x \quad \Longleftrightarrow \\ M_Ay = & \quad \lambda y \quad \text{where } y = Sx \neq 0. \end{aligned}$$

We note that the transformation from the second to third line follows from the fact that $L_Ae = 0$. These equations are all equivalent, so λ is an eigenvalue of both L_A and M_A for some eigenvector of L_A besides e. Therefore the eigenvalues of M_A must be the same as the eigenvalues of L_A sans the zero eigenvalue; additionally, the eigenvectors of M_A must be of the form S applied to the eigenvectors of L_A. We shall now show that the non-diagonal entries of M_A are non-positive. It must be the case that $(SL_A)_{i,j} = -l_{i,j} + l_{i+1,j}$ for all i, j, so we can show that

$$m_{i,j} = \sum_{k=1}^{n} (SL_A)_{i,k} T_{k,j} = \sum_{k=j+1}^{n} (-l_{i,k} + l_{i+1,k}) = \sum_{k=j+1}^{n} (a_{i,k} - a_{i+1,k}).$$

As A is Robinson, $a_{i,k} \leq a_{i+1,k}$ for $i < k+1$ and so $m_{i,j} \leq 0$ for $i < j$. To show the same condition holds for $i > j$, we use the fact that $\sum_{j=1}^{n} l_{i,j} = 0$ to obtain

$$m_{i,j} = \sum_{k=j+1}^{n} (-l_{i,k} + l_{i+1,k}) = \sum_{k=1}^{j} (-a_{i,k} + a_{i+1,k}).$$

This implies that $m_{i,j} \leq 0$ for $i > j$ due to the fact that A is Robinson. Therefore, every non-diagonal entry of M is non-positive. Now, we shall let β be greater then $\max_i\{\lambda_i, m_{i,i}\}$, where λ_i are the eigenvalues of M_A. Then the auxiliary matrix $\hat{M}_A := \beta I - M_A$ is nonnegative with eigenvalues $\hat{\lambda}_i = \beta - \lambda_i$ and $\hat{M}_A$ shares the same eigenvectors with M_A. By the Perron-Frobenius theorem, there exists a nonnegative eigenvector y of $\hat{M}_A$ corresponding to the smallest eigenvalue of M_A. However, y is also an eigenvector of M_A corresponding to the smallest eigenvalue of M_A; this is Sx, where x is the Fiedler vector of L_A. Since $y = Sx$ is nonnegative, the corresponding Fiedler vector is monotonic and the original claim is true. □

Proposition 2.1.13 (Theorem 3.3, [4]). *Suppose A is a Robinsonian matrix with a simple Fiedler value whose Fiedler vector x has no repeated values. If Π is a permutation matrix such that Πx is strictly monotonic, then $\Pi A \Pi^T$ is a Robinson matrix.*

Proof. As the Fiedler value is simple, its corresponding Fiedler vector is unique up to a multiplicative constant. Furthermore, if x is the Fiedler vector for A, then Πx is the Fiedler vector for $\Pi A \Pi^T$. Thus, applying a permutation matrix simultaneously to the rows and columns of A will simply permute the entries of the Fiedler vector. Now, we suppose that Π^* is a permutation matrix such that $\Pi^* A (\Pi^*)^T$ is a Robinson matrix.

By Lemma 2.1.12, $\Pi^* x$ is monotone, and because x has no repeated values, $\Pi^* x$ must be either x arranged in increasing or decreasing order. □

This proposition provides the basis for the spectral algorithm, though in this form it is too restrictive, as it requires the Fiedler vector to be simple and contain no repeated values. This can be handled with the following two observations: If A is Robinsonian, then $A + \alpha I$ is Robinsonian for any $\alpha \in \mathbb{R}$. Thus, it is acceptable to assume that the starting matrix at the beginning of the algorithm has smallest off-diagonal entry of 0. Furthermore, if A is reducible, then the problem of finding a Robinson ordering can be decoupled; finding a Robinson ordering for each of its irreducible blocks and concatenating them together will form a Robinson ordering for A. This is stated more formally as follows.

Lemma 2.1.14 (Lemma 4.2, [4]). *Let A_i with $1 \leq i \leq k$ be the irreducible blocks of a Robinsonian matrix A, and let Π_i be a permutation of the block A_i such that $\Pi_i A_i (\Pi_i)^T$ is Robinson. Then any permutation formed by concatenating the Π_i will be a Robinson ordering for A.*

These ideas, combined with a few lemmas, can be used to show the following, which solves the issue of the Fiedler vector of A not being simple.

Proposition 2.1.15 (Theorem 4.6, [4]). *If A is an irreducible Robinsonian matrix with $A_{n,1} = 0$, then the Fiedler value λ_2 is a simple eigenvalue.*

All that is left is to handle the repeated values case, which can be shown to decouple the problem into pieces that can be solved recursively. We state this more formally below.

Proposition 2.1.16 (Theorem 4.7, [4]). *Let A be a Robinsonian matrix with a simple Fiedler value and Fiedler vector x. Suppose there is a repeated value β in x; de ne $I, J, K \subseteq \{1, \ldots, n\}$ such that*

$$x_i < \beta \text{ for all } i \in I$$

$$x_i = \beta \text{ for all } i \in J$$

$$x_i > \beta \text{ for all } i \in K.$$

Then Π induces a Robinson ordering for A if and only if Π or Π^T can be expressed as (Π_i, Π_j, Π_k), where Π_j induces a Robinson ordering on the submatrix $A(J, J)$ of A induced by J, and Π_i, Π_k are the restrictions of a permutation matrix that induces a Robinson ordering on A to I and K respectively.

Thus the algorithm can be broken down into the following key steps, which we present here. Given a matrix A,

- Shift all entries of the matrix A so that it becomes nonnegative.
- Sort the matrix into its irreducible blocks $\{A_1, \ldots, A_k\}$. If there are multiple blocks, perform this algorithm on each block. Otherwise, proceed to next step.
- Compute the Fiedler vector x of L_A and then sort it so that x becomes monotone. For the jth distinct value of the sorted x, perform this algorithm on the submatrix $A(V_j, V_j)$, where V_j is the set of indices where x is equal to that jth distinct value.
- Use the newly acquired orderings of $A(V_j, V_j)$ to tiebreak the order of x. Then the sorting of x obtained is the Robinson ordering of A.

In [32], the authors use this algorithm—slightly modified for their purpose—to attack the *ranking problem*, where n items are given with pairwise relationships between them and the goal is to derive a total order on the items consistent with those relationships. They do so by developing a novel way to express a list of binary relations between items as a similarity matrix, then applying this algorithm. They also show that it is resistant to noise (in the ℓ^2 norm) when the noise is sufficiently random; as long as the errors caused by noise are not too numerous, the proper ordering will be recovered. They also discuss that $\sqrt{n}$ more comparisons are needed to recover the proper ordering in the case of noise in the ℓ^∞ scheme, showing its vulnerability to errors.

2.1.2 Approximation of Robinson orderings

It is often the case that matrices recovered from practical applications suffer from missing data or noisy interference; how can hidden patterns be discovered in a matrix when the matrix itself has been obscured? The algorithms discussed in Section 2.1.1 can perfectly recover Robinson orderings for matrices that possess them, but only in the entirely noiseless case. To deal with this problem, a new approach is needed—rather than find an ordering that is guaranteed to make a given matrix Robinson, the goal becomes to find a Robinson matrix such that the distance between this matrix and the given original is acceptably small in some norm. We shall discuss results proven about this problem in order of descending size of norm, starting with the ℓ^∞ norm.

Given a symmetric matrix A and a fixed $\epsilon > 0$, we refer to finding a Robinson matrix R such that $\|A - R\|_\infty < \epsilon$ as an ***ϵ-Robinson ℓ^∞- tting.*** It is not always the case that such a fitting exists; it was shown in [17] that determining if a fitting exists for a pair (A, ϵ) is NP-complete. They also show that the following associated problem

$$\begin{aligned} \text{Minimize} \quad & \|A - R\|_\infty \\ \text{Subject to} \quad & R \text{ is Robinsonian} \end{aligned} \tag{2.1}$$

is NP-hard to approximate with a factor smaller than 3/2. As a reminder, a polynomial time algorithm is an ***α-factor approximation*** algorithm for a minimization problem Π if for each instance of I of Π, it returns a solution whose value is at most α times the optimal value $\mathrm{OPT}_\Pi(I)$ of Π on I plus some independent constant. For more on factor approximations we refer the reader to [86]. To show that finding an ϵ-Robinson ℓ^∞-fitting is NP-complete, the authors of [18] construct a polynomial transformation from the NP-complete problem NOT-ALL-EQUAL 3-SAT; additionally, they show that if Problem (2.1) can be approximated by a factor less than 3/2, such an algorithm could solve the NOT-ALL-EQUAL 3-SAT problem in polynomial time. As this is impossible unless P=NP, this problem is thus considered to be NP-hard.

The same authors develop a factor-16 algorithm for the problem (2.1) in [18] that runs in $O(n^6 \log n)$ time. They begin with relaxing the idea of a Robinson ordering: Given $\epsilon \geq 0$ and a symmetric matrix A, a total order $\prec$ on the index set of A is ***ϵ-compatible*** if $u \prec x \prec y \prec v$ implies that $A_{u,v} + 2\epsilon \geq A_{x,y}$. They then show that the optimal error ϵ^* lies in a well-defined list $\mathcal{L}$ of size $O(n^4)$ and that for a given total order $\prec$, a Robinsonian matrix R compatible with $\prec$ with best fitting A can be found in $O(n^2)$. We note that this does not violate the claim that (2.1) is NP-hard, as this optimality is only for one fixed ordering; perhaps there is a different ordering $\prec'$ such that the Robinson minimizer R' results in a smaller ℓ^∞ distance with A. The algorithm then tests the entries of $\mathcal{L}$ using a parameter ϵ which represents the "guess" of the optimal error. For $\epsilon \in \mathcal{L}$, the algorithm will either find that no ϵ-compatible order exists, in which case $\|A - R\|_\infty$ for Robinsonian R, or it returns an ordering that is 16ϵ-compatible.

To describe the algorithm in more detail, we have to introduce some definitions from [18].

Definition 2.1.17 (Holes in a partial order)**.** Let $\epsilon > 0$. Furthermore, let $\prec$ be a total order and $\preccurlyeq$ be a partial order of a set $A = \{a_1, \ldots, a_n\}$. Select some maximal chain $P = (a_1, \ldots, a_m)$ of $\preccurlyeq$; we say that two consecutive elements $a_i, a_{i+1} \in P$ form a ***hole*** H_i and that all elements $x \in A \setminus P$ assigned between a_i and a_{i+1} are ***located*** in H_i.

This assignment of elements in $A \setminus P$ is done manually; we describe its process below. For an element $x \in A \setminus P$, define $a_x = \max\{a_i \in P \mid a_i \preccurlyeq x\}$ and $b_x = \min\{a_i \in P \mid x \preccurlyeq a_i\}$. Let a_x^* be the element after a_x in the chain P; similarly, let b_x^* be the element before b_x in the chain P. Denote by $H(x)$ the union of all holes $H_i = [a_i, a_{i+1}]$ comprised between a_x and b_x—such that we have $a_x \preccurlyeq a_i \preccurlyeq a_{i+1} \preccurlyeq b_x$—and call $H(x)$ the ***segment*** of x. The holes $H_x = [a_x, a_x^*]$ and $H^x = [b_x^*, b_x]$ are called the ***bounding holes*** of $H(x)$ and all other holes are called the ***inner holes***.

Additionally, define $I(x) = \{a_i \in P \mid a_x \preccurlyeq a_i \preccurlyeq b_x\}$ and $I'(x) = I(x) \setminus \{a_x, b_x\}$. The term A_x will be defined as follows:

$$A_x := \frac{1}{2}\Bigg(\min(A_{x,a_i} \mid a_i \in I'(x)) + \max(A_{x,a_i} \mid a_i \in I'(x)) \Bigg).$$

We note that as $x \notin P$, the segment $H(x)$ must contain at least two holes, implying that the bounding holes must be different. With this background in mind, we introduce the following definition.

Definition 2.1.18 (Admissible and pairwise admissible holes)**.** Let $x \in A \setminus P$. A hole H_i of $H(x)$ is called ***x-admissible*** if the total order on $P \cup \{x\}$ obtained from 4 by adding the relation a_i 4 x 4 a_{i+1} is ϵ-compatible with A. We note that the bounding holes of $H(x)$ must always be x-admissible. A pair $\{H_i, H_j\}$ of holes is called (x, y, c)***-admissible*** if H_i is x-admissible, H_j is y-admissible, and the total order on $P \cup \{x, y\}$ obtained from 4 by adding the relations a_i 4 x 4 a_{i+1} and a_j 4 y 4 a_{j+1} is $c\epsilon$-compatible with A.

We will denote by $AH(x)$ the set of all x-admissible holes H_i such that for each $y \in X \setminus P$ distinct from x there exists a y-admissible hole H_j where $\{H_i, H_j\}$ is an $(x, y, 1)$-admissible pair. Furthermore, it is not difficult to show that the bounding holes of $H(x)$ must themselves belong to $AH(x)$. For two vertices a_i, a_j with $i < j$ in the chain P, we denote by H_{ij} the union of all holes comprised between a_i and a_j. We then let X_{ij} be the set of all $x \in A \setminus P$ such that $H(x) = H_{ij}$.

The authors also introduce the following directed graphs:

Definition 2.1.19 (Hole distance graphs)**.** For each set X_{ij}, define a directed graph $\mathcal{L}_{ij} = (X_{ij}, E_{ij})$ as follows: $x \to y$ if either of the following conditions hold

- $A_{x,y} < \max(A_x, A_y) + 5\epsilon$
- x and y must be located in the same hole and for any ϵ-compatible order $\prec$ we must have either $a_{i+1} \prec x \prec y$ or $y \prec x \prec a_{j-1}$.

Another directed graph $\mathcal{G}_{ij} = (\mathcal{V}_{ij}, \mathcal{E}_{ij})$ is defined as follows with the intent to display the "big picture" behavior of $\mathcal{L}_{ij}$. The vertices $\mathcal{V}_{ij}$ are the strongly connected components of $\mathcal{L}_{ij}$ and an arc between two vertices C and C' is drawn if either of the following conditions holds:

- There exists an element x belonging to C and an element y belonging to C' such that $A_{x,y} > \max(A_x, A_y) + 3\epsilon$.
- There exist two elements x, x' belonging to C and two elements y, y' belonging to C' such that the pairs xx' and yy' are strongly linked and we have that $x \to y$, $x' \to y'$.

We now finally describe the approximation algorithm given in [18].

1. Construct the list $\mathcal{L}$ of feasible values for the optimal error ϵ^*. For each $\epsilon \in \mathcal{L}$, construct the canonical partial order $\preccurlyeq$ and compute a maximal chain $P = a_1 \preccurlyeq a_2 \preccurlyeq \ldots \preccurlyeq a_p$ of $(X, \preccurlyeq)$.

2. For each element $x \in X \setminus P$, compute the sets $AH(x)$ and $H(x)$. Then, for each pair $i < j - 1$, construct the set X_{ij} and make a bipartition $X_{ij} = \{X_{ij}^-, X_{ij}^+\}$, where all elements of X_{ij}^- are located in hole H_i and all elements of X_{ij}^+ are located in hole H_{j-1}.

3. While these two partition elements cannot be totally ordered, the vertices of $\mathcal{G}_{ij}^-$ and $\mathcal{G}_{ij}^+$ can be; compute those total orderings and denote them $\preccurlyeq_{ij}^-$ and $\preccurlyeq_{ij}^+$ respectively.

4. Concatenate in a single total order on these vertices these restricted total orders that come from different sets assigned to the same hole.

5. Recursively apply this process to each set that comprises a vertex of any $\mathcal{G}_{ij}^-$ or $\mathcal{G}_{ij}^+$; the returned total orders are concatenated into a single total order $\prec$ on X according to the total orders between vertices and between holes.

6. The order $\prec$ given by the smallest $\epsilon \in \mathcal{L}$ such that the algorithm does not halt is selected and returned as the output.

It is shown in [18, Section 5] that this algorithm must terminate in $O(n^6 \log n)$ time, and as a factor-16 algorithm its result must have error at most 16 times the actual optimal error for (2.1). A problem for future interested researchers could be improving the factor of this approximation, either with more sophisticated techniques, new ideas, or both.

Expanding our worldview to look at how this problem might fare for the ℓ^p norm, for Robinson matrices, the author has struggled to find results. However, for *strongly Robinson matrices*, we need look no further than [6, Theorem 2], which shows that this approximation problem is NP-complete for all $1 \leq p < \infty$. We begin with a definition.

Definition 2.1.20 (Strongly Robinson matrix)**.** Let A be a symmetric matrix. We say that A is *strongly Robinson* if for any indices $x < y < z < t$, the following conditions hold:

(i) $A_{x,z} \leq \min(A_{x,y}, A_{y,z})$

(ii) $A_{x,z} = A_{y,z}$ implies that $A_{x,t} = A_{y,t}$.

(iii) $A_{y,t} = A_{y,z}$ implies that $A_{x,z} = A_{x,t}$.

Let A be a given symmetric matrix. The following problem:

$$\begin{array}{rl} \text{Minimize} & \|A - R\|_p \\ \text{Subject to} & R \text{ is strongly Robinson} \end{array}$$

is shown in [6] to be NP-complete by considering a binary version of this problem, where both A and R must be $0 - 1$ valued. This alternate problem, which is clearly less complex, is then shown to be equivalent to the problem of finding a Hamiltonian path in a given graph G. As this problem is well known to be NP-complete, it must be so that the binary strongly Robinson approximation problem is NP-complete, and therefore so is the general version.

We have seen so far that finding the best ordering to approximate norm difference is a challenging question—what if the ordering was given, and the challenge became to find the best Robinson approximation of a fixed matrix? It was shown in [18] that given an ordering $\prec$ and a symmetric matrix A indexed by some set X ordered by $\prec$, it is possible to find a Robinson matrix R such that $\|A - R\|_\infty$ is minimized in $O(n^2)$ time. In fact, it is relatively simple to state the form of the minimizing matrix. Define a new matrix r_A as follows:

$$(r_A)_{i,j} := \max\{A_{u,v} \mid u, v \in X \text{ and } i \prec u \prec v \prec j\}.$$

Setting $\epsilon_A = \frac{1}{2}\|A - r_A\|_\infty$ and $\delta = \min\{A_{x,y} \mid x \neq y\}$, we define the new matrix R_A as

$$(R_A)_{i,j} := \max\{(r_A)_{i,j} - \epsilon_A, \delta\}.$$

Chepoi and Seston show in [18, Lemma 2.4] that R_A achieves the minimal value of $\|A - R_A\|_\infty$ while being Robinson, and it is clear that calculating R_A would be of time complexity $O(n^2)$. This is a far cry from the NP-hard problem that is the arbitrary ordering case. A natural follow up question is to ask how these results extend to the ℓ^p norms, if they do at all; however, this author is unaware of results studied for $1 < p < \infty$. For the other extreme case where $p = 1$, while the results known are not as exact, the question can be formed as a linear program (see [40]): Minimize the linear function $\|A - R\|_1$ subject to the constraint that R is Robinson. As this constraint can be expressed in $O(n^3)$ inequalities, the problem can be solved in polynomial time. We discuss further results on ℓ^1 approximations, as well as results for approximation with a smaller norm, more in Section 4.1.

2.2 Dense graph limits

The notion of dense graph limits was introduced by Lovász and Szegedy in [67]; here, a novel method for measuring the similarity of graphs was developed alongside a notion of convergence for sequences of graphs possessing said similarity. In this section, we will develop the necessary background required for understanding the use of this concept throughout the thesis, beginning with one of the key observations that galvanized Lovász and Szegedy's work: Graphs can be identified with symmetric functions on the unit square.

2.2.1 Kernels and graphons

In order to define the notion of convergence of dense graph sequences in any meaningful way, there are first quite a few things that must be made clear. We shall start by defining the basic objects upon which this premise relies and work our way forward from that starting point. A useful measure of the similarity between two graphs is that of determining whether they share subgraphs, induced or no. An exceedingly useful way to quantify this is to introduce the notion of homomorphism densities, to be defined shortly, as we will see that they lead to a rather natural definition of convergence for graph sequences.

Definition 2.2.1 (Homomorphism density)**.** Given two simple graphs F and G, the ***homomorphism density*** $t(F,G)$ is the probability that a random mapping from $V(F)$ to $V(G)$ is a homomorphism (an edge-preserving map). One may also realize this quantity as

$$t(F,G) = \frac{\hom(F,G)}{|V(G)|^{|V(F)|}},$$

where $\hom(F,G)$ is the number of homomorphisms from F into G. The quantity $t_0(F,G)$ is the probability that a random injection from $V(F)$ to $V(G)$ is a homomorphism; it is referred to as the ***injective homomorphism density*** and is similarly defined as

$$t_0(F,G) = \frac{\mathrm{inj}(F,G)}{|V(G)|(|V(G)|-1)...(|V(G)|-|V(F)|+1)},$$

where $\mathrm{inj}(F,G)$ is the number of injective homomorphisms from F into G. The ***induced homomorphism density*** $t_{ind}(F,G)$ is the probability that a random embedding of F into G as an induced subgraph is a homomorphism. If $\mathrm{ind}(F,G)$ is the number of such maps from F into G, then

$$t_{ind}(F,G) = \frac{\mathrm{ind}(F,G)}{|V(G)|(|V(G)|-1)...(|V(G)|-|V(F)|+1)}.$$

We also find that it is almost always more worthwhile to work with functions instead of graphs (functions respond well to standard analytical tools, whereas graphs do not), and so we immediately generalize the notion of graphs to live in some function space and never look back.

Definition 2.2.2 (Graphon space)**.** Let $\mathcal{W}$ denote the space of all bounded, symmetric, measurable functions $w : [0,1]^2 \to \mathsf{R}$. We shall refer to the elements of $\mathcal{W}$ as ***kernels***. Most of the time, one need only consider those elements $w \in \mathcal{W}$ such that $0 \leq w \leq 1$. We shall refer to this subset of $\mathcal{W}$ as $\mathcal{W}_0$ and refer to its elements as ***graphons***.

We note that kernels generalize weighted graphs in the following way.

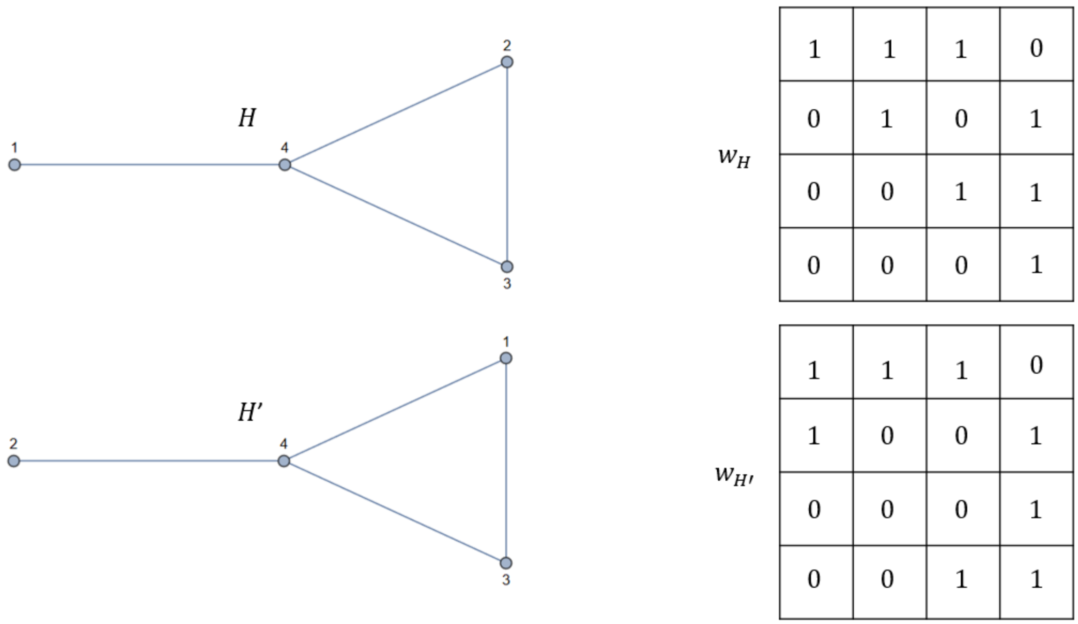

Figure 2.2: Two isomorphic labelled graphs H and H' with different associated step functions. Note that w_H and $w_{H'}$ are symmetric as functions on $[0,1]^2$, not as matrices.

Definition 2.2.3 (Step function)**.** A function $w \in \mathcal{W}$ is called a ***step function*** if there is a partition of $[0,1]$ into measurable sets $S_1 \cup ... \cup S_k$ such that w is constant on the sets $S_i \times S_j$. The sets S_i are the ***steps*** of w.

For every weighted graph H on nodeset $V(H) = \{1, ..., n\}$, with nodeweights α_i and edgeweights β_{ij}, one can define a step function $w_H \in \mathcal{W}$ in the following way. Divide $[0,1]$ into n intervals $I_1, ..., I_n$ such that $|I_i| = \alpha_i/\alpha_H$ (here α_H is the sum of all nodeweights of H), and for $x \in I_i$ and $y \in I_j$, let $w_H(x,y) = \beta_{ij}$. Importantly, note that w_H is dependent on the labeling of the graph H, and is therefore not invariant under graph isomorphisms. We refer to Figure 2.2 for a visual example. Conversely, every step function u with steps of equal size gives rise to a weighted graph: If $S_1, ..., S_k$ are its steps, then the graph has node set $\{1, ..., k\}$ and the edge ij has weight $u(x,y)$, where $x \in S_i$ and $y \in S_j$.

As one should hope, homomorphism densities in graphs extend to homomorphism densities in graphons and kernels. For every $w \in \mathcal{W}$ and every simple graph $F = (V, E)$, define

$$t(F, w) = \int_{[0,1]^{|V(F)|}} \prod_{ij \in E} w(x_i, x_j) \prod_{i \in V} dx_i.$$

One can attempt to visualize this by picturing $[0, 1]$ as the nodeset of a graph, where the edge between vertices x and y is given weight $w(x, y)$. It holds that for every simple graph F and every weighted graph G,

$$t(F, G) = t(F, w_G),$$

because we have that

$$\begin{aligned} t(F, w_G) &= \int_{[0,1]^{|V(F)|}} \prod_{ij \in E(F)} w(x_i, x_j) \prod_{i \in V(F)} dx_i \\ &= \int_{[0,1]^{|V(F)|}} \prod_{ij \in E(F)} \mathbb{1}(ij \in E(G)) \mathbb{1}(I_i \times I_j) \prod_{i \in V(F)} dx_i \\ &= \frac{\hom(F, G)}{|V(G)|^{|V(F)|}} = t(F, G). \end{aligned}$$

This relation shows the use of this choice of extension. We note also that this equality does not depend on the labeling of G; this is because changes in G are reflected in w_G.

2.2.2 The cut distance

Now that we have a function space, we introduce a norm. Specifically, we shall work with the ***cut norm***, which was introduced by [38] and is defined below.

Definition 2.2.4 (Cut norm)**.** Let $w \in \mathcal{W}$. We define the ***cut norm*** by

$$\|w\| \quad = \sup_{S,T \subseteq [0,1]} \left| \iint_{S \times T} w(x, y) dx dy \right|$$

where the supremum is taken over all measurable subsets S and T. We define the cut norm of a graph G simply by $\|G\| := \|w_G\|$.

It is obvious that for any function $w \in \mathcal{W}$,

$$\|w\| \leq \|w\|_1 \leq \|w\|_p \leq \|w\|_\infty$$

for all $p \geq 1$. We note that any kernel defined on the fixed set $[0,1]$ corresponds to a labeled graph. If we wish to look at so-called "unlabeled" kernels, we must introduce another notion of distance.

Definition 2.2.5 (Cut distance)**.** Let Φ be the set of all measure preserving bijections from $[0,1]$ to $[0,1]$; that is, if $A \subset [0,1]$, then $|\Phi(A)| = |A|$. The ***cut distance*** between two kernels u and w is

$$\delta(u,w) = \inf_{\phi\in\Phi} \|u - w^\phi\|,$$

where $w^\phi(x,y) = w(\phi(x), \phi(y))$.

One can think of the maps ϕ as relabelings of the set $[0,1]$, and thus we take the infimum over them to see whether there is any labeling where the difference between the two kernels becomes zero. It is a generalization of checking whether there exists an isomorphism between two graphs, though we note it is not a direct analogue, as two graphs with different nodesets can have cut distance zero. Thus the distance δ is only a pseudometric; we can identify kernels of cut distance 0 to get the set $\widetilde{\mathcal{W}}$ of ***unlabeled kernels***. $\widetilde{\mathcal{W}_0}$ is defined similarly.

2.2.3 The operator T_w

While the cut norm sees most of its use in a combinatorial setting, it is equivalent to other more traditional norms, such as the operator norm of T_w where T_w is an operator $L_\infty \to L_1$. In this case, as well as in other cases, the operator T_w takes the form

$$(T_w f)(x) = \int_0^1 w(x,y) f(y) dy.$$

One can simply check that for all $w \in \mathcal{W}$,

$$\|w\| \quad \leq \|T_w\|_{\infty \to 1} \leq 4\|w\| \quad .$$

Note that we can also think of T_w as an operator from $L_1 \to L_\infty$ or from $L_2 \to L_2$; in either case, the definition remains the same, but in the latter case the operator becomes one of considerably more interest. $T_w : L_2 \to L_2$ is what is known as a ***Hilbert-Schmidt operator*** (for more details, see [76]); i.e. it is a bounded linear operator on a Hilbert space such that

$$\sum_{i \in I} \|T_w e_i\|^2 < \infty$$

where $\|\cdot\|$ is the norm of the Hilbert space (in our case, the standard L_2 norm) and the e_i form some orthonormal basis for the Hilbert space. It does not matter which basis is chosen, as the sum is independent of such choice. Hilbert-Schmidt operators, like anything bearing the name of one or more famous mathematicians, have very nice properties. First and foremost among those properties is their compactness; they have a countable multiset $\mathrm{Spec}(w)$ of eigenvalues $\{\lambda_1, \lambda_2 ...\}$ such that $\lambda_n \to 0$. In particular, this implies that every nonzero eigenvalue has finite multiplicity.

With the introduction of T_w, we are now able to prove that the cut norm and cut distance do not just exist as extrema—for the cut norm, the supremum is achieved, and likewise for the cut distance, the infimum is achieved.

Lemma 2.2.6. *[66, Lemma 8.10] For any graphon $w \in \mathcal{W}$, the values*

$$\sup_{S,T \subset [0,1]} \left| \iint_{S \times T} w \right| \quad \textit{and} \quad \sup_{f,g:[0,1] \to [0,1]} \left| \iint_{[0,1]^2} f(x)g(y)w(x,y)dxdy \right|$$

are attained, and they are both equal to $\|w\|$. Here the functions f, g and the sets S, T are assumed to be measurable.

Proof. Let $C = \sup_{f,g} \langle f, T_w g \rangle$ and let $f_n, g_n : [0,1] \to [0,1]$ for $n = 1, 2, \ldots$ be measurable functions such that $\langle f_n, T_w g_n \rangle \to C$. As the set of measurable functions from $[0,1]$

to itself is weak*-compact, there must then exist a subsequence f_{n_k} that converges to f in the sense that $\langle f_n, \phi \rangle \to \langle f, \phi \rangle$ for all $\phi \in L^1[0,1]$. A further subsequence can be selected to show that g_n converges to g in a similar sense. We now claim that

$$\iint_{[0,1]^2} f_n(x) g_n(y) w(x,y) dx dy \to \iint_{[0,1]^2} f(x) g(y) w(x,y) dx dy.$$

This convergence is trivial for the case when $w = \mathbb{1}_{S \times T}$ for measurable sets $S, T \subseteq [0,1]$. It must then follow when w is a step function, as step functions are linear combinations of functions of the aforementioned type. From this the claim follows for all graphons $w \in \mathcal{W}$, as every graphon can be approximated by step functions in $L^1([0,1]^2)$ and the functions f_n, g_n, f, and g are bounded. Thus, $\langle f, T_w g \rangle = C$.

The next step is to show that these functions f and g can be chosen to be $0-1$ valued. Let $S = \{x : 0 < f(x) < 1\}$ and suppose that $|S| > 0$. Define

$$f_s(x) = f(x) + s \min(f(x), 1 - f(x)).$$

For $-1 \leq s \leq 1$, the function f_s satisfies $0 \leq f_s \leq 1$ and thus by the maximality of f it must be that $\langle f_s, T_w g \rangle \leq \langle f, T_w g \rangle$. Because $\langle f_s, T_w g \rangle$ is linear in s and equality holds for $s = 0$, equality must hold for all values of s. In particular, it must hold for $s = 1$, which means we can replace f by $f_1(x) = \min(1, 2f(x))$. Repeating this construction, we create a sequence of optimizing functions that monotone converges to the $0-1$ valued function $\overline{f} = \mathbb{1}_{f(x)>0}$. f can therefore be replaced by $\overline{f}$, and a similar argument can be made to replace g with a $0-1$ valued function $\overline{g}$, proving the initial claim. □

We show also in the case of step functions that the set achieving the cut norm must be a disjoint union of the function's steps.

Lemma 2.2.7. *Let $u \in \mathcal{W}$ be a step function on the steps $\{S_1, \ldots, S_k\}$ and let the measurable sets S, T be the sets that achieve $\|u\|$. Then $S = \bigcup_{i \in I} S_i$ and $T = \bigcup_{j \in J} S_j$ for some index sets I and J.*

Proof. Let u be a step function defined on the steps $\{S_1, \dots, S_k\}$ that partition $[0,1]$. Specifically, let $u = \sum_{i,j} a_{i,j} \mathbb{1}_{S_i \times S_j}$, where $a_{i,j} \in \mathsf{R}$. By Lemma 2.2.6, there exist measurable sets $S, T \subseteq [0,1]$ such that, without loss of generality, $\|u\| = \iint_{S\times T} u$. Therefore,

$$\|u\| = \iint_{S\times T} u = \sum_{i,j=1}^{k} \iint_{(S\cap S_i)\times(T\cap S_j)} a_{i,j} = \sum_{i,j=1}^{k} a_{i,j} |S\cap S_i||T\cap S_j|.$$

Let i be fixed. We then split into two cases; in the first case, the quantity $\sum_{j=1}^{k} a_{i,j}|T\cap S_j| \geq 0$. Here,

$$\sum_{j=1}^{k} a_{i,j}|T\cap S_j||S_i| \geq \sum_{j=1}^{k} a_{i,j}|T\cap S_j||S\cap S_i|,$$

which by the maximality of $\|u\|$ forces $|S\cap S_i| = |S_i|$. In the second case, the quantity $\sum_{j=1}^{k} a_{i,j}|T\cap S_j| < 0$. Here,

$$0 > \sum_{j=1}^{k} a_{i,j}|T\cap S_j||S\cap S_i|,$$

which by the maximality of $\|u\|$ forces $|S\cap S_i| = 0$. Thus, $S\cap S_i$ must either be the empty set or S_i. An identical argument can be performed to establish the corresponding result for T. □

2.2.4 The stepping operator

In the business of graphons, one will commonly have to deal with the average of functions over cells, as is evidenced by the norm that encodes convergence of graphs/graphons being concerned entirely with finding the cell with the maximal average. Thus, some notation is introduced to allow us to easily refer to common objects related to cell averages.

Definition 2.2.8 (Stepping operator)**.** Let $w \in \mathcal{W}$ and let $\mathcal{P} = (S_1, ..., S_k)$ where we assume $|S_i| > 0$ be a partition of $[0,1]$ into a finite number of nonempty measurable sets. We define the function $w_{\mathcal{P}}$ by

$$w_{\mathcal{P}}(x,y) = \frac{1}{|S_i||S_j|} \iint_{S_i \times S_j} w(x,y)dxdy \quad (x \in S_i, y \in S_j),$$

where the operator $w \mapsto w_{\mathcal{P}}$ is called the ***stepping operator***.

Essentially, $w_{\mathcal{P}}$ is obtained by averaging w over the cells $S_i \times S_j$; it is a step function whose steps are $\mathcal{P}$. This construction is called a ***stepping*** of w. We note that if we consider the stepping operator on the space $\mathcal{W}$, it is a linear operator that is both idempotent and symmetric; i.e., for all $u, w \in \mathcal{W}$,

$$\langle u_{\mathcal{P}}, w_{\mathcal{P}} \rangle = \langle u, w_{\mathcal{P}} \rangle = \langle u_{\mathcal{P}}, w \rangle \tag{2.2}$$

where $\langle f, g \rangle = \iint_{[0,1]^2} fg \, dxdy$. Another important property of the stepping operator in regards to the p-norm and the cut norm is shown in the following lemma. We note that this property is sometimes referred to as being ***contractive***, as the stepping operator can never increase the p-norm or cut norm of a graphon w and can only maintain or decrease it.

Lemma 2.2.9. *Let $w \in \mathcal{W}$, $\mathcal{P}$ be any partition of $[0,1]$ into a nite number of nonempty measurable sets, and let $p \geq 1$; then the following is true.*

i.) $\|w_{\mathcal{P}}\|_p \leq \|w\|_p$.

ii.) $\|w_{\mathcal{P}}\| \leq \|w\|$.

Proof. Let $w \in \mathcal{W}$ and let $\mathcal{P} = \{S_1, ..., S_n\}$ be a partition of $[0,1]$ into n measurable sets. We note that for $p \geq 1$,

$$\begin{aligned}
\|w_{\mathcal{P}}\|_p &= \left(\iint_{[0,1]^2} |w_{\mathcal{P}}|^p dxdy \right)^{\frac{1}{p}} = \left(\sum_{i,j} |S_i \times S_j| \left| \iint_{S_i \times S_j} \frac{w}{|S_i \times S_j|} dxdy \right|^p \right)^{\frac{1}{p}} \\
&= \left(\sum_{i,j} |S_i \times S_j|^{1-p} \left| \iint_{S_i \times S_j} w \; dxdy \right|^p \right)^{\frac{1}{p}} \leq \left(\sum_{i,j} \frac{|S_i \times S_j|^{1-p}}{|S_i \times S_j|^{-\frac{p}{q}}} \iint_{S_i \times S_j} |w|^p dxdy \right)^{\frac{1}{p}} \\
&= \left(\iint_{[0,1]^2} |w|^p dxdy \right)^{\frac{1}{p}} = \|w\|_p,
\end{aligned}$$

where $\frac{1}{p} + \frac{1}{q} = 1$ and the inequality is due to Hölder's inequality. Thus it is clear that $i.$) holds. To prove $ii.$), let $S \times T$ be the set which achieves $\|w_{\mathcal{P}}\|_\square$. By Lemma 2.2.7 $S = \bigcup_{i \in I} S_i$ and $T = \bigcup_{j \in J} S_j$ for some $I, J \subset \{1, \ldots, n\}$; thus we can show that

$$\begin{aligned} \|w_{\mathcal{P}}\|_\square &= \left| \iint_{S \times T} w_{\mathcal{P}} \right| = \left| \sum_{i,j} \iint_{S_i \times \mathcal{S}_j} w \frac{|S_i \cap S||S_j \cap T|}{|S_i||S_j|} \right| \\ &= \left| \sum_{\substack{i \in I \\ j \in J}} \iint_{S_i \times \mathcal{S}_j} w \right| = \left| \iint_{S \times T} w \right| \\ &\leq \sup_{S,T} \left| \iint_{S \times T} w \right| = \|w\|_\square , \end{aligned}$$

where the supremum is taken over measurable subsets of $[0, 1]$, completing the proof.

□

2.2.5 Compactness of graphon space

The Regularity Lemma of Szemerédi [85] and its extensions are among the most important tools in understanding the behavior of large dense graphs; naturally, one hopes it has a clean extension to graphons. As it turns out, it does indeed, and we state now the Weak Regularity Lemma for kernels.

Lemma 2.2.10 ([66], Lemma 9.9)**.** *For every function $w \in \mathcal{W}$ and $k \geq 1$, there is a step function u with k steps such that*

$$\|w - u\|_\square < \frac{2}{\sqrt{\log k}} \|w\|_2.$$

This may seem trivial—it is well-known that step functions are dense in L_1 and so any kernel w can be arbitrarily approximated by step functions in the L_1-norm (and thus the cut norm). However, the beauty of this lemma comes from the fact that the error bound depends only upon the number of steps used in the step function u. We do note, however, that these steps need not be intervals or the same size. In order to prove this result, we first must state and prove a different lemma

Lemma 2.2.11 ([66], Lemma 9.11)**.** *Let $w \in \mathcal{W}$ and let $\mathcal{P}$ be a measurable k-partition of $[0,1]$. Then the following must be true.*

i.) There are two sets $S, T \subseteq [0,1]$ and a real number $0 \le a \le \|w\|_\infty$ such that

$$\|w - a\mathbf{1}_{S\times T}\|_2^2 \le \|w\|_2^2 - \|u\|^2 .$$

ii.) There is a partition $\mathcal{Q}$ re ning $\mathcal{P}$ with at most $4k$ classes such that

$$\|w - w_{\mathcal{P}}\| \;\; = \|w_{\mathcal{Q}} - w_{\mathcal{P}}\| \;\; .$$

We note that this lemma implies that for any kernel $w \in \mathcal{W}$ and $k \ge 1$, there are k pairs of subsets $S_i, T_i \subseteq [0,1]$ which need not be disjoint and k real numbers a_i such that

$$\left\| w - \sum_{i=1}^{k} a_i \mathbf{1}_{S_i\times T_i} \right\| \;\; < \frac{1}{\sqrt{k}}. \tag{2.3}$$

To show this, we apply the above lemma repeatedly to get pairs of sets S_i, T_i and real numbers a_i such that the functions $w_j = w - \sum_{i=1}^{j} a_i \mathbf{1}_{S_i\times T_i}$ satisfy

$$\|w_j\|_2^2 \le \|w\|_2^2 - \sum_{i=0}^{j-1} \|w_i\|^2$$

and note that since the right hand must always be nonnegative, for every k there exists $0 \le i < k$ with $\|w_i\|^2 \;\le \frac{1}{k}$. Simply set $a_{i+1}, ..., a_k$ to 0 to get (2.3). It is also clear that the function $\sum_i a_i \mathbf{1}_{S_i\times T_i}$ is a stepfunction and that it can be made symmetric by taking its average with $\sum_{i=1}^{k} a_i \mathbf{1}_{T_i\times S_i}$, which will give a total of $2k$ terms; one can check that this will have at most 2^{2k} steps, and so 2.2.10 follows from (2.3). We now present the proof of Lemma 2.2.11.

Proof of Lemma 2.2.11. Let S, T be measurable subsets of $[0,1]$ such that

$$\|w\| \;\; = \left| \iint_{S\times T} u \right| = |\langle w, \mathbf{1}_{S\times T}\rangle|,$$

where we assume that $\langle w, \mathbf{1}_{S\times T}\rangle \geq 0$. Let $a = \frac{1}{|S||T|}\|w\|$. Then

$$\begin{aligned}
\|w - a\mathbf{1}_{S\times T}\|_2^2 &= \langle w - a\mathbf{1}_{S\times T}, w - a\mathbf{1}_{S\times T}\rangle \\
&= \langle w, w\rangle - \langle w, a\mathbf{1}_{S\times T}\rangle - \langle a\mathbf{1}_{S\times T}, w\rangle + \langle a\mathbf{1}_{S\times T}, a\mathbf{1}_{S\times T}\rangle \\
&= \|w\|_2^2 - 2a\|w\| \ + a^2|S||T| = \|w\|_2^2 - \frac{1}{|S||T|}\|w\|^2 \\
&\leq \|w\|_2^2 - \|w\|^2 .
\end{aligned}$$

This proves i.). Proving ii.) is similar. We get $\|w - w_{\mathcal{P}}\| \ \geq \|w_{\mathcal{Q}} - w_{\mathcal{P}}\|$ from the contractive property of the stepping operator; to prove the other needed inequality, we let S, T be measurable subsets of $[0,1]$ such that $|\langle w - w_{\mathcal{P}}, \mathbf{1}_{S\times T}\rangle| = \|w - w_{\mathcal{P}}\|$, and let $\mathcal{Q}$ denote the partition generated by $\mathcal{P}, S$, and T. It is clear that $\mathcal{Q}$ has at most $4k$ classes. By (2.2), it must be that $\langle w, \mathbf{1}_{S\times T}\rangle = \langle w_{\mathcal{P}}, \mathbf{1}_{S\times T}\rangle$, and thus

$$\|w - w_{\mathcal{P}}\| \ = |\langle w - w_{\mathcal{P}}, \mathbf{1}_{S\times T}\rangle| = |\langle w_{\mathcal{Q}} - w_{\mathcal{P}}, \mathbf{1}_{S\times T}\rangle| \leq \|w_{\mathcal{Q}} - w_{\mathcal{P}}\| \ .$$

This completes the proof. □

In [67], Lovász and Szegedy proved a theorem that is both equivalent to all versions of the Regularity Lemma and also of tremendous practical use. We shall recall its statement and proof below.

Theorem 2.2.12 ([66], Theorem 9.23)**.** *The space* $(\widetilde{W}_0, \delta \ )$ *is compact.*

Proof. It suffices to prove that every sequence of graphons $w_1, w_2, ...$ has a convergent subsequence. To do this, we will create subsequences $\{w_{n_k}\}$ of $\{w_n\}$ such that the steppings of $\{w_{n_k}\}$ (with respect to carefully chosen partitions) will converge to functions $\{u_k\}$, who themselves form a martingale and thus converge to a limit object u. We then show that there must exist a subsequence of $\{w_n\}$ (without steppings this time) that converges to u, proving compactness. We start with the following: For every $n \geq 1$, we can construct partitions $\mathcal{P}_{n,k}$ of $[0,1]$ such that these partitions and the corresponding step functions $w_{n,k} := (w_n)_{\mathcal{P}_{n,k}}$ satisfy the following conditions:

i.) $\|w_n - w_{n,k}\|_\square \le \frac{1}{k}$,

ii.) The partition $\mathcal{P}_{n,k+1}$ refines $\mathcal{P}_{n,k}$,

iii.) $|\mathcal{P}_{n,k}| = m_k$ depends only on k.

Once we have these partitions, we can rearrange the points of $[0,1]$ by a measure preserving bijection so that every partition class in every $\mathcal{P}_{n,k}$ is an interval. We also claim that we can replace the sequence (w_n) by a subsequence so that for every k, the sequence $w_{n,k}$ converges almost everywhere to a stepfunction u_k with m_k steps as $n \to \infty$.

This is done by selecting a subsequence of the w_n for which the length of the i-th interval of $\mathcal{P}_{n,1}$ converges for every i as $n \to \infty$ (note that this is possible as the number of intervals in $\mathcal{P}_{n,1}$ is fixed), and the value of $w_{n,1}$ on the product of the i-th and j-th intervals converges for every i and j (as $n \to \infty$). It then follows that the sequence $w_{n,1}$ must converge almost everywhere to a limit object u_1, which itself must be a step function with m_1 steps that are intervals.

We repeat this process for $k = 2, 3, ...$ to get subsequences for which $w_{n,k} \to u_k$ almost everywhere, where u_k is a stepfunction with m_k steps that are intervals. This provides a subsequence with the desired properties. We let $\mathcal{P}_k$ denote the partition of $[0,1]$ into the steps of u_k. For every $k < l$, the partition $\mathcal{P}_{n,l}$ is a refinement of the partition $\mathcal{P}_{n,k}$, and thus $w_{n,k} = (w_{n,l})_{\mathcal{P}_{n,k}}$. It is simple to see that this is inherited by the limiting stepfunctions:

$$u_k = (u_l)_{\mathcal{P}_k}.$$

This implies that if (X,Y) is a random point in $[0,1]^2$ chosen uniformly, then the sequence $(u_1(X,Y),\, u_2(X,Y), ...)$ must be a martingale. Because the random variables $u_i(X,Y)$ are bounded, the Martingale Convergence Theorem ([87], Theorem 9.23) implies that the sequence is convergent with probability 1. Thus, the sequence of functions $(u_1, u_2, ...)$ is convergent almost everywhere; we call its limit u. We will show that $\delta_\square(u, w_n) \to 0$ as $n \to \infty$.

Fix any $\epsilon > 0$. Then there must exist a $k > \frac{3}{\epsilon}$ such that $\|u - u_k\|_1 < \frac{\epsilon}{3}$. If this k is fixed, there must also exist an n_0 such that $\|u_k - w_{n,k}\|_1 < \frac{\epsilon}{3}$ for all $n \geq n_0$. We note also that $\delta\ \ (w_{n,k}, w_n) \leq \|w_{n,k} - w_n\|\ \ \leq \frac{1}{k} < \frac{\epsilon}{3}$. Thus

$$\begin{aligned}
\delta\ \ (u, w_n) &\leq \delta\ \ (u, u_k) + \delta\ \ (u_k, w_{n,k}) + \delta\ \ (w_{n,k}, w_n) \\
&\leq \|u - u_k\|_1 + \|u_k - w_{n,k}\|_1 + \delta\ \ (w_{n,k}, w_n) \\
&< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.
\end{aligned}$$

This completes the proof. □

We note that this theorem remains valid if $\widetilde{\mathcal{W}}_0$ is replaced by ***any*** uniformly bounded subset of $\widetilde{\mathcal{W}}$ that is closed in the $\delta\ \ $ distance.

2.2.6 Sampling

Every graphon w gives rise to a rather natural method of generating random graphs; when we talk of "sampling" from a graphon, we are talking about creating one of these random graphs. Given a graphon w and a set $S = \{x_1, ..., x_n\}$ where each $x_i \in [0, 1]$, one can create a weighted graph $\mathcal{H}(S, w)$ on the node set $[n]$ by assigning the weight $w(x_i, x_j)$ to edge ij, where $i, j \in [n]$ and $i \neq j$.

Every weighted graph H (with edgeweights $\beta_{ij}(H) \in [0, 1]$) gives rise to a random simple graph $\mathcal{G}(H)$ on $V(H)$: Connect vertices i and j with probability $\beta_{ij}(H)$, making independent decisions for distinct pairs (i, j), where $i, j \in [n]$ and $i \neq j$.

We consider the particular random graph $\mathcal{G}(S, w) = \mathcal{G}(\mathcal{H}(S, w))$. For some integer n, we define the random weighted graph $\mathcal{H}(n, w) = \mathcal{H}(S, w)$ and the random simple graph $\mathcal{G}(n, w) = \mathcal{G}(S, w)$, where S is an ordered n-tuple of independent uniform random points on $[0, 1]$. We refer to $\mathcal{G}(n, w)$ as a ***w-random graph***. Importantly, for large enough samples, the cut distance between a graphon and a sample is small with high probability.

Lemma 2.2.13 ([66], Lemma 10.16)**.** *Let* $k \geq 1$ *and let* $w \in \mathcal{W}_0$ *be a graphon. Then with probability at least* $1 - \exp(-k/2 \log k)$*,*

$$\delta\ \ (\mathcal{H}(k, w), w) \leq \frac{20}{\sqrt{\log k}}$$

and

$$\delta\ \ (\mathcal{G}(k, w), w) \leq \frac{22}{\sqrt{\log k}}.$$

We refer to [66] for the proof. We can now relate the two fundamental quantities that we use to study dense graphs and graphons: the cut distance and homomorphism densities $t(F, \cdot)$. The following simple relationship is due to Lovász and Szegedy and is a generalization of the Counting Lemma in the theory of Szemerédi partitions.

Lemma 2.2.14. *Let* F *be a simple graph and let* $w, w' \in \mathcal{W}_0$*. Then*

$$|t(F, w) - t(F, w')| \leq |E(F)|\delta\ \ (w, w').$$

We will actually prove this lemma in the more general setting of $\mathcal{W}_0$-decorated graphs, as it simplifies the proof. A $\mathcal{W}_0$-decorated graph is a simple graph in which a graphon w_e is assigned to each edge e. For every $\mathcal{W}_0$-decorated graph (F, w) we define

$$t(F, w) = \int_{[0,1]^{|V(F)|}} \prod_{ij \in E(F)} w_{ij}(x_i, x_j) \prod_{i \in V(F)} dx_i.$$

Lemma 2.2.15 ([66], Lemma 10.24)**.** *Let* (F, w) *and* (F, w') *be two* $\mathcal{W}_0$*-decorated graphs with the same underlying simple graph, where* $w = (w_e : e \in E)$ *and* $w' = (w'_e : e \in E)$*. Then*

$$|t(F, w) - t(F, w')| \leq \sum_{e \in E} \|w_e - w'_e\|\ \ .$$

This suffices to prove the claim of Lemma 2.2.14, as $f(F, w') = t(F, (w')^{\phi})$ for any measure preserving bijection $\phi : [0,1] \to [0,1]$; thus, for any such ϕ, we have that

$$|t(F,w) - t(F,w')| = |t(F,w) - t(F,(w')^{\phi}| \leq \sum_{e \in E} \|w_e - (w'_e)^{\phi}\| \ . \tag{2.4}$$

Therefore, we can take two $\mathcal{W}_0$-decorated graphs (F,w) and (F,w') with the same underlying simple graph F such that (F,w) has every edge decorated by the same graphon w and (F,w') has every edge decorated by the same graphon w'. Then, taking the infimum over all such ϕ in on both sides of (2.4), we achieve

$$|t(F,w) - t(F,w')| \leq |E(F)|\delta \ (w,w'),$$

which is precisely the statement of Lemma 2.2.14.

Proof. It suffices to prove this statement for the case when $w_e = w'_e$ for all edges but one. To see why, if multiple edges differed, this same argument would work and result in a sum of cut norm differences. Let $F = (V,E)$, and let uv be the edge with $w_{uv} \neq w'_{uv}$. Then

$$\begin{aligned} t(F,w) - t(F,w') &= \int_{[0,1]^{|V(F)|}} \prod_{ij \in E \setminus \{uv\}} w_{ij}(x_i,x_j)(w_{uv}(x_u,x_v) - w'_{uv}(x_u,x_v))dx \\ &= \int_{[0,1]^{|V(F)|}} f(x)g(x)(w_{uv}(x_u,x_v) - w'_{uv}(x_u,x_v))dx \end{aligned}$$

where

$$f(x) = \prod_{ij \in \nabla(u) \setminus \{uv\}} w_{ij}(x_i,x_j)$$

($\nabla(u)$ is the set of edges incident to u) does not depend on x_v and satisfies $0 \leq f \leq 1$. Similarly,

$$g(x) = \prod_{ij \in E \setminus \nabla(u)} w_{ij}(x_i,x_j)$$

does not depend on x_u and satisfies $0 \leq g \leq 1$. Fixing all variables except x_u and x_v, we get

$$\left| \iint_{[0,1]^2} f(x)g(x)(w_{uv}(x_u,x_v) - w'_{uv}(x_u,x_v))dx_u dx_v \right| \leq \|w_{uv} - w'_{uv}\| \ \ .$$

Integrating over the remaining variables gives

$$|t(F,w) - t(F,w')| \leq \|w_{uv} - w'_{uv}\| \ \ ,$$

proving the result. □

Therefore convergence in homomorphism densities forces convergence in the cut distance; amazingly, the converse also holds true. This is typically known as the "Inverse Counting Lemma", and shows that convergence in $\delta\ \ $ implies convergence of homomorphism densities, albeit at a much slower rate than the previous lemma. We note that it is typical to measure the distance between samples using the cut norm, but oftentimes (such as in the upcoming proof) it is useful to consider the ***variation distance***: If α and β are probability measures on the Borel space $(\Omega, \mathcal{F})$, then

$$d_{var}(\alpha,\beta) = \sup_{\sigma \in \mathcal{F}} |\alpha(\sigma) - \beta(\sigma)|.$$

Informally, this is the largest possible difference between the probabilities that the two probability measures can assign to the same event. For more information, refer to [41].

Lemma 2.2.16 ([66], Lemma 10.32)**.** *Let k be a positive integer, let $u, w \in \mathcal{W}_0$, and assume that for every simple graph F on k nodes,*

$$|t(F,u) - t(F,w)| \leq 2^{-k^2}.$$

Then

$$\delta\ \ (u,w) \leq \frac{50}{\sqrt{\log k}}.$$

Proof. Assume that $u, w \in \mathcal{W}_0$ satisfy

$$|t(F,u) - t(F,w)| \leq 2^{-k^2}$$

for every simple graph F with k nodes. By inclusion-exclusion, this implies that

$$|t_{ind}(F,u) - t_{ind}(F,w)| \leq 2^{\binom{k}{2}} 2^{-k^2} = 2^{-\binom{k+1}{2}}.$$

Thus in terms of the w-random graphs $\mathcal{G}(k,u)$ and $\mathcal{G}(k,w)$,

$$|\mathsf{P}(\mathcal{G}(k,u) = F) - \mathsf{P}(\mathcal{G}(k,w) = F)| \leq 2^{-\binom{k+1}{2}},$$

which implies that

$$\begin{aligned} d_{var}(\mathcal{G}(k,u), \mathcal{G}(k,w)) &= \sum_F |\mathsf{P}(\mathcal{G}(k,u) = F) - \mathsf{P}(\mathcal{G}(k,w) = F)| \\ &\leq 2^{\binom{k}{2}} 2^{-\binom{k+1}{2}} = 2^{-k} \leq 1 - 2\exp\left(-\frac{k}{2\log k}\right). \end{aligned}$$

This implies that we can couple $\mathcal{G}(k,u)$ and $\mathcal{G}(k,w)$ so that $\mathcal{G}(k,u) = \mathcal{G}(k,w)$ with probability larger than $2\exp(-\frac{k}{2\log k})$, which by Lemma 10.16 of [66] implies that

$$\delta\ \ (u, \mathcal{G}(k,u)) \leq \frac{22}{\sqrt{\log k}}$$

with probability at least $1 - 2\exp(-\frac{k}{2\log k})$, where similar results hold for w. Thus with positive probability all three events occur and so

$$\delta\ \ (u,w) \leq \delta\ \ (u, \mathcal{G}(k,u)) + \delta\ \ (w, \mathcal{G}(k,w)) \leq \frac{50}{\sqrt{\log k}},$$

proving our initial claim. □

2.2.7 Convergence

We define now—finally—what it means for a sequence of dense graphs $\{G_n\}$ to be convergent. While there are several ways of defining convergence for dense graphs, the most used and most studied version is that involving homomorphism densities, which we present below.

Definition 2.2.17 (Convergence of graph sequences)**.** A sequence of graphs $\{G_n\}$ is ***convergent*** if and only if for every simple graph F, $\{t(F, G_n)\}$ is a convergent sequence of real numbers.

Because $t(F, G_n) = t(F, w_{G_n})$, we shall consider the convergence of the sequences $\{t(F, w_{G_n})\}$ instead. This notion of convergence is referred to as ***left convergence***, as it arises based on homomorphisms "from the left". In the case of dense graphs it is usually referred to simply as convergence although in the sparse case that will be introduced in a later section there are multiple methods of convergence that abound in use. However, the idea of basing this convergence around each homomorphism density of a graph seems clunky; it would be impractical to check that every such sequence of densities converges for any graph of import. Therefore, throughout the thesis we have introduced and strengthened a rather robust tool for this very purpose—the cut distance δ .

Theorem 2.2.18. *A sequence $\{G_n\}$ of simple graphs with $|V(G_n)| \to \infty$ as $n \to \infty$ is convergent if and only if it is a Cauchy sequence in the metric δ .*

Proof. The Counting Lemma (2.2.14) implies that every Cauchy sequence in the metric δ is convergent, implying the sequence $t(F, w_{G_n})$ must be Cauchy, which forces convergence. The Inverse Counting Lemma (2.2.16) applied to w_{G_n} implies the converse. □

This proof builds on a fair amount of previous work conducted over the last few pages and does the important job of proving the equivalence of left convergence and convergence in δ . Thus we now are free to consider convergence of dense graph sequences in whichever framework we like (hint: it will almost always be δ). We now also go over convergence in more practical detail. It is known from [67] that every convergent sequence of dense graphs has a limit graphon and that every graphon is the limit of a convergent sequence of dense graphs. We shall prove the latter; it follows from the statement below.

Theorem 2.2.19 ([67]). *Let F be a simple graph on k vertices. Then for every $0 < \epsilon < 1$,*

$$\mathsf{P}(|t(F, \mathcal{G}(n, w)) - t(F, w)| > \epsilon) \leq 2 \exp\Big(- \frac{\epsilon^2}{18k^2} n\Big),$$

where n is the desired number of vertices in $\mathcal{G}(n, w)$.

The sum of the right hand side is convergent for every fixed $\epsilon, k > 0$, so it follows from the Borel-Cantelli Lemma that $t(F, \mathcal{G}(n, w)) \to t(F, w)$ with probability 1. Then, because there are only a countable number of graphs F, this statement holds with probability 1 for every graph F, showing convergence of $\{\mathcal{G}(n, w)\}$ in n.

Proof of Theorem 2.2.19. We proceed by forming a martingale in the following way: In the m-th step, we create a new vertex and $X_m \in [0, 1]$, then generate the edges of $\mathcal{G}$ connecting the new node to previously generated nodes based on X_m. The probability that a random injection of $V(F)$ into $V(G)$ is a homomorphism, conditioning on the parts of the graph already generated, is a martingale. Applying Azuma's Inequality [5] to this martingale achieves the desired result.

Let $k = |V(F)|$, $\mathcal{G} = \mathcal{G}(n, w)$, and $\mathcal{G}_m$ be the subgraph of $\mathcal{G}$ induced by the vertices $1, ..., m$. For every injection $\phi : [k] \to [m]$, let A_ϕ denote the event that ϕ is a homomorphism from F to the random graph $\mathcal{G}$. Define

$$B_m = \frac{1}{n(n-1)...(n-k+1)} \sum_{\phi} \mathsf{P}(A_\phi | \mathcal{G}_m).$$

Note the following.

$$\begin{aligned}\mathsf{E}(B_m|\mathsf{G}_{m-1}) &= \mathsf{E}\left(\frac{1}{(n)_k}\sum_{\phi}\mathsf{P}(A_\phi|\mathsf{G}_m)|\mathsf{G}_{m-1}\right) = \frac{1}{(n)_k}\sum_{\phi}\mathsf{E}(\mathsf{E}(\mathbb{1}_{A_\phi}|\mathsf{G}_m)|\mathsf{G}_{m-1})\\ &= \frac{1}{(n)_k}\sum_{\phi}\mathsf{E}(\mathbb{1}_{A_\phi}|\mathsf{G}_{m-1}) = \frac{1}{(n)_k}\sum_{\phi}\mathsf{P}(A_\phi|\mathsf{G}_{m-1}) = B_{m-1}.\end{aligned}$$

Thus $(B_0, B_1, ...)$ is a martingale. Furthermore, it is clear that $\mathsf{P}(A_\phi) = t(F, w)$ and that

$$\mathsf{P}(A_\phi|\mathcal{G}_n) = \begin{cases} 1 & \text{if } \phi \text{ is a homomorphism from } F \text{ to } G,\\ 0 & \text{if otherwise.}\end{cases}$$

Therefore we can note that

$$B_0 = \sum_{\phi}\mathsf{P}(A_\phi) = t(F, w),$$

and

$$B_n = \frac{1}{n(n-1)...(n-k+1)}\mathrm{inj}(F, \mathcal{G}) = t_0(F, \mathcal{G}).$$

Hence we must find bounds on $|B_m - B_{m-1}|$, which is done through the all too familiar triangle inequality. Consider

$$\begin{aligned}|B_m - B_{m-1}| &= \frac{1}{n(n-1)...(n-k+1)}\left|\sum_{\phi}(\mathsf{P}(A_\phi|\mathsf{G}_m) - \mathsf{P}(A_\phi|\mathsf{G}_{m-1}))\right|\\ &\le \frac{1}{n(n-1)...(n-k+1)}\sum_{\phi}\left|\mathsf{P}(A_\phi|\mathsf{G}_m) - \mathsf{P}(A_\phi|\mathsf{G}_{m-1})\right|.\end{aligned}$$

In this sum, every term for which m is not in the range of ϕ is 0, and the other terms are at most 1. The number of terms of the latter kind is $k(n-1)...(n-k+1)$, and so

$$|B_m - B_{m-1}| \le \frac{k(n-1)...(n-k+1)}{n(n-1)...(n-k+1)} = \frac{k}{n}.$$

Azuma's Inequality [5] then shows that we have

$$\mathsf{P}(|B_n - B_0| > \epsilon) = \mathsf{P}(|t(F,w) - t_0(F,\mathcal{G})| > \epsilon) \le \exp\Big(-\frac{\epsilon^2}{2k^2}n\Big).$$

We note that for $n > \frac{k^2}{\epsilon}$ we have that

$$|t_0(F,\mathcal{G}) - t(F,\mathcal{G})| \le \frac{\epsilon}{3},$$

and so (2.2.19) follows by replacing ϵ with $\frac{\epsilon}{3}$ in the above proof. □

2.3 Sparse graph limits

While the above theory provides deep and meaningful results about many important families of graphs, it does so only for sequences of dense graphs. A sequence of graphs $\{G_n\}$ is called ***sparse*** if $\|w_{G_n}\|_1 \to 0$ as $n \to \infty$; these are the graphs that are not dense. The above theory mandates that G_n must converge to the identically zero graphon; for any simple graph F with at least one edge, $t(F, G_n) \to 0$. This is uninteresting and unhelpful; surely there must be another way to define a useful mode of convergence for sparse graphs? In [11], Bollobas and Riordan developed a method of convergence for very sparse graphs—graphs with uniformly bounded maximum (average) degree–that shows graphons remain the proper limiting object in such cases. However, the boundedness assumption they make is highly restrictive; many networks of interest (such as those whose edge sets are governed by a power law regime) are sparse but do not satisfy this condition.

This is resolved in [12], where Borgs, Chayes, Cohn, and Zhao define a convergence method that provides nontrivial results for convergence of sparse graph sequences without enforcing a degree boundedness condition. This is done primarily by normalizing the graphs by their edge density, or in terms of the associated graphons, normalizing with respect to their L^1 norm. This will naturally require an extension of the previous theory to unbounded functions on $[0,1]^2$ that are not bounded; we shall present that here, along with the novel form of convergence.

2.3.1 L^p graphons

As with dense graph limit theory, we will work mainly with graphons and kernels. However, as mentioned above, these functions do not need to be bounded. We shall work with so-called L^p graphons, defined below.

Definition 2.3.1 (L^p graphons)**.** Let $p \geq 1$. An L^p ***graphon*** is a symmetric, measurable function $w \in L^p([0,1]^2)$. We refer to the space of L^p graphons as $\mathcal{W}^p$.

We note that by Hölder's Inequality, for $p < q$, $\mathcal{W}^q \subseteq \mathcal{W}^p$. We will still be working with $\|\cdot\|$, but we will also be dealing with the more familiar $\|\cdot\|_p$ as well, defined as it usually is.

2.3.2 L^p upper regularity

Just as all sequences of numbers are not convergent, neither are all sequences of L^p graphons. A sequence of L^p graphons converging in δ does not imply its limit object is in L^p. To guarantee this, we must define a new concept called L^p upper regularity, introduced in [12].

Definition 2.3.2 (L^p upper regularity for graphons)**.** A graphon $w \in \mathcal{W}^p$ is (C,η)-***upper L^p regular*** if whenever $\mathcal{P}$ is a partition of $[0,1]$ into measurable sets each having measure at least η,

$$\|w_{\mathcal{P}}\|_p \leq C.$$

For an associated graphon w_G, we can think of this as saying that for every partition of the vertices in which no part is too small, the weighted graph derived from averaging the edge weights with respect to the partition is bounded in the L^p norm. This notion naturally extends to general L^p graphons as shown above. We note that enforcing a uniform upper bound on the L^p norm of the associated graphon itself is actually too strong; it would correspond to a lower bound on $\|w_G\|_1$, forcing the graphon to be dense. As with the dense theory, we prefer to think of graphs in terms of their associated graphons, and thus present the next definition.

Definition 2.3.3 (L^p upper regularity for graphs)**.** A simple graph G is (C,η)-*upper L^p regular* if the normalized associated graphon $w_G/\|w_G\|_1$ is (C,η)-upper L^p regular.

We introduce a final definition involving L^p upper regularity that defines the graph/graphon sequences of interest in the sparse theory.

Definition 2.3.4 (L^p upper regular sequences)**.** Let $1 < p \leq \infty$ and $C > 0$. We say $\{w_n\}$ is a C-*upper L^p regular sequence* of graphons if for every $\eta > 0$ there is some $n_0 \in \mathsf{N}$ such that w_n is $(C+\eta,\eta)$-upper L^p regular for all $n \geq n_0$. An L^p upper regular sequence of graphs is defined similarly.

A natural question about sparse convergence is how the sequence of p-norms $\{\|w_n\|_p\}_{n\in\mathsf{N}}$ behaves when a sequence $\{w_n\} \subset \mathcal{W}^p$ is convergent in δ . The following example shows that w_n converging (in cut norm) to w in the space $\mathcal{W}^p$ does not imply convergence of the associated p-norms—as mentioned before, L^p upper regularity is necessary.

Example 2.3.5. For fixed $1 < p < \infty$, consider the random graph model G_n on n vertices given by connecting vertices i and j with probability $\min(1, n^{\frac{1}{p+1}}(ij)^{-\frac{1}{p+1}})$. It is known (see [12, Introduction]) that the edge density of G_n is of the order of $n^{-\frac{1}{p+1}}$ and thus tends to 0, but that when G_n is rescaled by its edge density, its δ -limit is $w(x,y) = (xy)^{-\frac{1}{p+1}}$. If we denote $w_n = \frac{1}{\|G_n\|_1} w_{G_n}$, then $\{w_n\}$ is a sequence of graphons converging to w, but $\|w_n\|_p$ is of the order of $n^{\frac{p-1}{p(p+1)}}$, which tends to infinity.

2.3.3 Sparse convergence

We can now state the main convergence results of [12], which assert that any sequence of L^p upper regular graphs/graphons must have at least a subsequential limit.

Theorem 2.3.6 (Theorem 2.8, [12])**.** *Let $p > 1$ and let $\{G_n\}$ be a C-upper L^p regular sequence of graphs. Then there exists an L^p graphon w with $\|w\|_p \leq C$ so that*

$$\liminf_{n\to\infty} \delta \left(\frac{G_n}{\|G_n\|_1}, w \right) = 0.$$

A result for graphons follows immediately.

Theorem 2.3.7 (Theorem 2.9, [12])**.** *Let $p > 1$ and let $\{w_n\}$ be a C-upper L^p regular sequence of graphons. Then there exists an L^p graphon w such that $\|w\|_p \leq C$ and*

$$\liminf_{n\to\infty} \delta \ (w_n, w) = 0.$$

These two limit results are proved in two parts: first, approximating of an L^p upper regular graph(on) by an L^p graphon with respect to the cut metric; second, establishing a limit result in the space of L^p graphons. The latter result of the two can be interpreted as a compactness result for L^p graphons analogous to the similar result shown for traditional graphons; we shall present the proof of both parts rather than prove the convergence results themselves. We begin with the approximation result, which can be viewed as a weak regularity lemma on the space of L^p graphons.

Proposition 2.3.8 (Proposition 2.17, [12])**.** *For every $p > 1$ and $\epsilon > 0$, there exists an $\eta > 0$ such that for every (C, η)-upper regular L^p weighted graph G (or graphon w), there exists an L^p graphon u with $\|u\|_p \leq C$ such that*

$$\delta \ \left(\frac{G}{\|G\|_1}, u\right) \leq C\epsilon \quad \textit{(respectively, } \delta \ (w, u) \leq C\epsilon).$$

Proof. We shall prove a result that implies Proposition 2.3.8 as a corollary; we proceed with the proof for the graphon case, as the notation is simpler. Let $C > 0$, $p > 1$, $0 < \epsilon < 1$, $N = (6/\epsilon)^{\max(2,p/(p-1))}$, and $\eta = 4^{-N-1}(\epsilon/160)^{p/(p-1)}$. Additionally, let w be a (C, η)-upper L^p regular graphon. Then we claim there exists a partition $\mathcal{P}$ of $[0, 1]$ into at most 4^N measurable parts, each having measure at least η, such that $\|w - w_{\mathcal{P}}\| \ \leq C\epsilon$.

To prove this, we consider a sequence of partitions $\mathcal{P}_0, \mathcal{P}_1, ..., \mathcal{P}_n$ of $[0, 1]$, beginning with the trivial partition $\mathcal{P}_0 = \{[0, 1]\}$. The following properties are enforced:

(i) The partition $\mathcal{P}_{i+1}$ refines $\mathcal{P}_i$ by dividing each part of $\mathcal{P}_i$ into at most four parts. This guarantees that $|\mathcal{P}_i| \leq 4^i$.

(ii) For each i, all parts of $\mathcal{P}_i$ have measure at least η.

The construction is as follows: For each $0 \le i < n$, if $\mathcal{P}_i$ satisfies $\|w - w_{\mathcal{P}_i}\|\ \ \le C\epsilon$, then we are done. Otherwise, there must exist measurable subsets $S, T \subseteq [0,1]$ with

$$|\langle w - w_{\mathcal{P}_i}, \mathbb{1}_{S\times T}\rangle| > C\epsilon. \tag{2.5}$$

Next, we find sets S', T' where both $|S\Delta S'|, |T\Delta T'| \le 2|\mathcal{P}_i|\eta$, such that if $\mathcal{P}_{i+1}$ is defined to be the common refinement of $\mathcal{P}_i, S'$, and T', then all parts of $\mathcal{P}_i$ have size at least η. To show that this is possible, we can look at the intersection of S with each part of $\mathcal{P}_i$ and obtain S' by deleting (rounding down) the parts that intersect with S in measure less than η and then adding (rounding up) the parts that intersect with S^C in measure less than η. Call this common refinement $\mathcal{P}_{i+1/2}$ such that all of its parts have measure at least η and $|S\Delta S'| \le |\mathcal{P}_i|\eta$. Now, we do a similar procedure to obtain T' so that the common refinement $\mathcal{P}_i$ of $\mathcal{P}_{i+1/2}$ and T has all parts with measure at least η and $|T\Delta T'| \le |\mathcal{P}_{i+1/2}|\eta \le 2|\mathcal{P}_i|\eta$, showing that $\mathcal{P}_i$ has the desired properties.

If the construction of these partitions stops at $\mathcal{P}_n$ with $n \le N$, we are done. Otherwise, we stop the construction at $\mathcal{P}_n$ with $n = \lceil N \rceil$ and work to show a contradiction. Let $0 \le i < n$ and let S, S', T, and T' be the sets used to construct $\mathcal{P}_{i+1}$ from $\mathcal{P}_i$. Then, using that $|S\Delta S'|, |T\Delta T'| \le 2|\mathcal{P}_i|\eta \le 2 \cdot 4^N\eta$, [12, Lemma 4.2] mandates that

$$|\langle w, \mathbb{1}_{S\times T} - \mathbb{1}_{S'\times T'}\rangle| \le 40C(2\cdot 4^N\eta)^{1-1/p} \le C\epsilon/4.$$

Furthermore, by Hölder's inequality (with $1/p + 1/p' = 1$)

$$\begin{aligned}|\langle w_{\mathcal{P}_i}, \mathbb{1}_{S\times T} - \mathbb{1}_{S'\times T'}\rangle| &\le \|w_{\mathcal{P}_i}\|_p \|\mathbb{1}_{S\times T} - \mathbb{1}_{S'\times T'}\|_{p'} \\ &\le C(|S\Delta S'| + |T\Delta T'|)^{1/p'} \\ &\le C(4\cdot 4^N\eta)^{1-1/p} \le C\epsilon/160 \le C\epsilon/8.\end{aligned}$$

It must then be true that

$$|\langle w - w_{\mathcal{P}_i}, \mathbb{1}_{S\times T}\rangle - \langle w - w_{\mathcal{P}_i}, \mathbb{1}_{S'\times T'}\rangle| \leq |\langle w - w_{\mathcal{P}_i}, \mathbb{1}_{S\times T}\rangle| + |\langle w - w_{\mathcal{P}_i}, \mathbb{1}_{S'\times T'}\rangle| \leq C\epsilon/2.$$

Combining this with (2.5) gives us that $|\langle w - w_{\mathcal{P}_i}, \mathbb{1}_{S'\times T'}\rangle| > C\epsilon/2$, and since S' and T' are both unions of parts in $\mathcal{P}_{i+1}$, we must also have that $\langle w, \mathbb{1}_{S'\times T'}\rangle = \langle w_{\mathcal{P}_{i+1}}, \mathbb{1}_{S'\times T'}\rangle$. This implies that

$$|\langle w_{\mathcal{P}_{i+1}} - w_{\mathcal{P}_i}, \mathbb{1}_{S'\times T'}\rangle| > C\epsilon/2, \tag{2.6}$$

and the proof now breaks into two parts.

Case 1: $p \geq 2$. Since $\mathcal{P}_{i+1}$ is a refinement of $\mathcal{P}_i$, we have that $\langle w_{\mathcal{P}_{i+1}} - w_{\mathcal{P}_i}, w_{\mathcal{P}_i}\rangle = 0$. So, by the Pythagorean theorem as well as the Cauchy-Schwarz inequality, we get that

$$\|w_{\mathcal{P}_{i+1}}\|_2^2 - \|w_{\mathcal{P}_i}\|_2^2 = \|w_{\mathcal{P}_{i+1}} - w_{\mathcal{P}_i}\|_2^2 \geq |\langle w - w_{\mathcal{P}_i}, \mathbb{1}_{S'\times T'}\rangle|^2 > C^2\epsilon^2/4.$$

This implies that $\|w_{\mathcal{P}_n}\|_2^2 > nC^2\epsilon^2/4 \geq NC^2\epsilon^2/4 > C^2$, which contradicts $\|w_{\mathcal{P}_n}\|_2 \leq \|w_{\mathcal{P}_n}\|_2 \leq C$.

Case 2: $1 < p < 2$. In this case, it is no longer as easy to form an upper bound on $\|w_{\mathcal{P}_n}\|_2$. Thus, we perform a truncation argument, stopping the partition construction at step n, truncating the last step function, and then calculating the energy that would have come from refining the truncated graphon. Set $K := C(6/\epsilon)^{1/(p-1)}$ and define the truncation $u := w_{\mathcal{P}_n}\mathbb{1}_{|w_{\mathcal{P}_n|\leq K}}$. We claim that for $0 \leq i < n$,

$$\|u_{\mathcal{P}_{i+1}}\|_2^2 > \|u_{\mathcal{P}_i}\|_2^2 + (C\epsilon/6)^2. \tag{2.7}$$

If this is the case, then $\|u_{\mathcal{P}_n}\|_2^2 > n(C\epsilon/6)^2 \geq N(C\epsilon/6)^2 = C^2(6/\epsilon)^{(2-p)/(p-1)}$, contradicting

$$\|u_{\mathcal{P}_n}\|_2^2 = \|w_{\mathcal{P}_n}\mathbb{1}_{|w_{\mathcal{P}_n|\leq K}}\|_2^2 \leq \|w_{\mathcal{P}_n}(K/|w_{\mathcal{P}_n}|^{1-p/2}\|_2^2$$

$$= \|w_{\mathcal{P}_n}\|_p^p K^{2-p} \le C^p K^{2-p} = C^2(6/\epsilon)^{(2-p)/(p-1)}.$$

We thus must only prove (2.7). To do so, we note that

$$\begin{aligned}
\|w_{\mathcal{P}_n} - u\|_1 &= \|w_{\mathcal{P}_n} \mathbb{1}_{|w_{\mathcal{P}_n}| \le K}\|_1 \\
&\le \|w_{\mathcal{P}_n}(|w_{\mathcal{P}_n}|/K)^{p-1}\|_1 \\
&= \||w_{\mathcal{P}_n}|^p\|_1/K^{p-1} = \|w_{\mathcal{P}_n}\|_p^p K^{p-1} \\
&\le C^p/K^{p-1} = C\epsilon/6.
\end{aligned}$$

Since $\mathcal{P}_n$ is a refinement of $\mathcal{P}_i$, it must be that $(w_{\mathcal{P}_n})_{\mathcal{P}_i} = w_{\mathcal{P}_i}$. Therefore,

$$\|w_{\mathcal{P}_i} - u_{\mathcal{P}_i}\|_1 = \|(w_{\mathcal{P}_n} - u)_{\mathcal{P}_i}\|_1 \le \|w_{\mathcal{P}_n} - u\|_1 \le C\epsilon/6. \tag{2.8}$$

Similarly, it must also be the case that $\|w_{\mathcal{P}_{i+1}} - u_{\mathcal{P}_{i+1}}\|_1 \le C\epsilon/6$. Using the triangle inequality, (2.6), and (2.8), we can show that

$$\begin{aligned}
|\langle u_{\mathcal{P}_{i+1}} - u_{\mathcal{P}_i}, \mathbb{1}_{S' \times T'}\rangle| \ge\ & |\langle w_{\mathcal{P}_{i+1}} - w_{\mathcal{P}_i}, \mathbb{1}_{S' \times T'}\rangle| \\
&- \|w_{\mathcal{P}_i} - u_{\mathcal{P}_i}\|_1 - \|w_{\mathcal{P}_{i+1}} - u_{\mathcal{P}_{i+1}}\|_1 \\
&> C(\epsilon/2 - \epsilon/6 - \epsilon/6) = C\epsilon/6.
\end{aligned}$$

Since $\mathcal{P}_{i+1}$ is a refinement of $\mathcal{P}_i$, we must have $\langle u_{\mathcal{P}_{i+1}} - u_{\mathcal{P}_i}, u_{\mathcal{P}_i}\rangle = 0$. So once more by the Pythagorean theorem and the Cauchy-Schwarz inequality, we get

$$\|u_{\mathcal{P}_{i+1}}\|_2^2 - \|u_{\mathcal{P}_i}\|_2^2 = \|u_{\mathcal{P}_{i+1}} - u_{\mathcal{P}_i}\|_2^2 \ge |\langle u_{\mathcal{P}_{i+1}} - u_{\mathcal{P}_i}, \mathbb{1}_{S' \times T'}\rangle|^2 > (C\epsilon/6)^2,$$

proving (2.7) as desired. □

We now present the compactness result, noting that L^1 is a special case where extra assumptions need to be made (namely uniform integrability) to ensure compactness and do not include its proof here.

Theorem 2.3.9 ([12]). *Let $1 < p \leq \infty$ and $C > 0$, and let $\{w_n\}_{n \geq 0}$ be a sequence of L^p graphons with $\|w_n\|_p \leq C$ for all n. Then there exists an L^p graphon w with $\|w\|_p \leq C$ such that*

$$\liminf_{n \to \infty} \delta \;\, (w_n, w) = 0.$$

Proof. Without loss of generality—scaling if necessary—we can assume that $C = 1$. For each k and n we construct an equipartition $\mathcal{P}_{n,k}$ using [12, Lemma 3.3] if $p \geq 2$ or [12, Lemma 3.4] when $1 < p < 2$ such that

$$\|w_n - (w_n)_{\mathcal{P}_{n,k}}\| \;\; \leq \frac{1}{k}.$$

Thus, we can assume that $\mathcal{P}_{n,k+1}$ always refines $\mathcal{P}_{n,k}$ and that $|\mathcal{P}_{n,k}|$ is independent of n. We now change variables so that each $\mathcal{P}_{n,k}$ become the same. We let $\mathcal{P}_k$ be a partition of $[0, 1]$ into $|\mathcal{P}_{n,k}|$ intervals of equal length, and for each n and k, we let $\psi_{n,k}$ be a measure preserving bijection from $[0, 1]$ to itself that transforms $\mathcal{P}_{n,k}$ into $\mathcal{P}$. Let

$$w_{n,k} := (w_n^{\psi_{n,k}})_{\mathcal{P}_k} = ((w_n)_{\mathcal{P}_{n,k}})^{\psi_{n,k}};$$

defined this way, $w_{n,k}$ is a step function with steps formed from $\mathcal{P}_k$ such that

$$\delta \;\, (w_n, w_{n,k}) \leq \frac{1}{k}.$$

Because each interval of $\mathcal{P}_k$ has length exactly $1/|\mathcal{P}_k|$ and the stepping operator is contractive with respect to the p-norm, we get that

$$|\mathcal{P}_k|^{-2} \|w_{n,k}\|_\infty^p \leq \|w_{n,k}\|_p^p \leq \|w_n\|_p^p \leq 1,$$

implying that $\|w_{n,k}\|_\infty \leq |\mathcal{P}_k|^{2/p}$. Next, we pass to a subsequence of $\{w_n\}$ such that for each k, we have that $w_{n,k}$ converges to a limit u_k a.e. as $n \to \infty$. For each fixed k, this done using the compactness of a $|\mathcal{P}_k|^2$-dimensional cube; this is because $w_{n,k}$ is determined by $|\mathcal{P}_k|^2$ steps and $\|w_{n,k}\|_\infty$ is uniformly bounded. To find a single subsequence that ensures convergence for all k, we iteratively choose a subsequence for $k = 1, 2, 3, \ldots$.

This sequence $u_1, u_2, \ldots$ was chosen because it forms a martingale on $[0,1]^2$ with respect to the σ-algebras generated by the products of $\mathcal{P}_1, \mathcal{P}_2, \ldots$, i.e. $(u_{k+1})_{\mathcal{P}_k} = u_k$. This can be shown immediately from

$$(w_{n,k+1})_{\mathcal{P}_k} = (w_n^{\psi_{n,k+1}})_{\mathcal{P}_k} = ((w_n)_{\mathcal{P}_{n,k}})^{\psi_{n,k+1}} = w_{n,k}.$$

By the L^p martingale convergence theorem [29, Theorem 5.4.5], there exists some $w \in L^p([0,1]^2)$ such that $\|u_k - w\|_p \to 0$ as $k \to \infty$. Since we know that $\|u_k\|_p \leq 1$ for all k, we must also have $\|w\|_p \leq 1$. Now we show that w is the desired limit by noting

$$\begin{aligned}\delta\ (w_n, w) &\leq \delta\ (w_n, w_{n,k}) + \delta\ (w_{n,k}, u_k) + \delta\ (u_k, w) \\ &\leq \delta\ (w_n, w_{n,k}) + \|w_{n,k} - u_k\|_1 + \|u_k - w\|_1.\end{aligned}$$

Each of these terms can be made aribitrarily small by choosing k and then n large enough. Thus, $\delta\ (w_n, w) \to 0$ along some subsequence and we are done. □

We now look at the other end of convergence; the behavior of a sequence of graphons $\{w_n\}$ (or graphs $\{G_n\}$) convergent to an L^p graphon w. The following proposition shows that if the behavior of the p-norms of the sequence is guaranteed, the sequence indeed must be L^p upper regular.

Proposition 2.3.10 (Proposition 2.10, [12]). *Let $1 \leq p \leq \infty$, let w be an L^p graphon, and let $\{w_n\}$ be a sequence of graphons ($\{G_n\}$ a sequence of graphs) such that*

$$\delta\ (w_n, w) \to 0 \quad \text{or} \quad \delta\left(\frac{G_n}{\|G_n\|}, w\right) \to 0$$

as $n \to \infty$. Then $\{w_n\}$ ($\{G_n\}$) is a $\|w\|_p$-upper regular L^p regular sequence.

Proof. We prove an auxiliary result from which this proposition follows. Let $C > 0$, $\eta > 0$, and $1 \leq p \leq \infty$, and let w be a (C, η)-upper L^p regular graphon. Let u

be another graphon. If $\|w-u\|\ \ \leq \eta^3$, then we claim u is $(C+\eta,\eta)$-upper regular L^p regular. To prove this, for any subsets $S,T \subseteq [0,1]$, we have $|\langle w-u, \mathbb{1}_{S\times T}\rangle| \leq \|w-u\|\ \ \leq \eta^3$. It then follows that

$$\left|\frac{1}{|S||T|}\left(\iint_{S\times T} w - \iint_{S\times T} u\right)\right| \leq \frac{\eta^3}{|S||T|} \leq \eta,$$

provided that $|S|,|T| \geq \eta$. Thus, for any partition $\mathcal{P}$ of $[0,1]$ into sets each having measure at least η we have that $|u_{\mathcal{P}} - w_{\mathcal{P}}| \leq \eta$ pointwise. Therefore,

$$\|u_{\mathcal{P}}\|_p \leq \||w_{\mathcal{P}}| + \eta\|_p \leq \|w_{\mathcal{P}}\|_p + \|\eta\|_p \leq C + \eta,$$

from which it follows that u is $(C+\eta,\eta)$-upper L^p regular. □

We must now show that there exists at least one sequence of sparse graphs that are convergent in the ways laid out in this section; we do so by developing a random graph model as in the dense theory, with the added restriction that the random graphs become sparse.

2.3.4 Sparse random graphs

The main result in this section is that every graphon w gives rise to a random graph model that produces a sequence of sparse graphs converging to w. These random models are made very similarly to their dense counterparts, with an extra parameter ρ added in. To make the graph $\mathcal{G}(n,w,\rho)$, simply let $\rho > 0$ and let $x_1,...,x_n$ be chosen independently and uniformly from $[0,1]$, then create an edge between vertices i and j with probability $\min(\rho|w(x_i,x_j)|,1)$. The random weighted graph $\mathcal{H}(n,w)$ is defined identically to the dense case. The main results for convergence of these graph models are showcased in the following theorem.

Theorem 2.3.11 (Theorem 2.14, [12])**.** *Let* $w \in \mathcal{W}^1$*. Then:*

i.) $\|\mathcal{H}(n,w) - w\|_1 \to 0$ *as* $n \to \infty$ *with probability 1.*

ii.) If ρ_n *satisfies* $\rho_n \to 0$ *and* $n\rho_n \to \infty$ *as* $n \to \infty$*, then*

$$\|\rho_n^{-1}\mathcal{G}(n,w,\rho_n) - w\| \quad \to 0$$

as $n \to \infty$ with probability 1.

As a simple corollary from the above theorem, for ρ_n satisfying the conditions above, $\mathcal{G}(n,w,\rho_n)$ must converge to w in the normalized cut metric with probability 1.

Proof of Theorem 2.3.11. For part *i.*), we begin by recalling a theorem of Hoeffding [46] that if w is a graphon and $x_1, x_2, ...$ a sequence of Uniform([0,1]) random variables uniformly chosen from $[0,1]$, then with probability 1:

$$\lim_{n\to\infty} \frac{1}{\binom{n}{2}} \sum_{1\le i<j\le n} w(x_i,x_j) \to \iint_{[0,1]^2} w(x,y)dxdy.$$

All weighted graphs $\mathcal{H}(\cdot,n)$ in this proof come from the same random sequence $x_1, x_2, ...$ of iid uniformly [0,1] variables. Fix $\epsilon > 0$. It suffices to show that $\limsup_{n\to\infty} \|\mathcal{H}(w,n) - w\|_1 \le \epsilon$ holds with probability 1. Let $\mathcal{P}$ denote the partition of $[0,1]$ into m equal intervals, where m is chosen to be sufficiently large such that $\|w - w_{\mathcal{P}}\|_1 \le \frac{\epsilon}{2}$. Fix this m and $\mathcal{P}$. Since the sequence $x_1, x_2, ...$ is equidistributed among the m intervals of $\mathcal{P}$,with probability 1 we have that $\|\mathcal{H}(w_{\mathcal{P}},n) - w_{\mathcal{P}}\|_1 \to 0$ as $n \to \infty$. We have $\|\mathcal{H}(w,n) - \mathcal{H}(w_{\mathcal{P}},n)\|_1 = \|H(w - w_{\mathcal{P}},n)\|_1$, which by (2.3.4) converges almost surely to $\|w - w_{\mathcal{P}}\|_1$. It therefore follows that with probability 1, the limit superior as $n \to \infty$ of

$$\|\mathcal{H}(w,n) - w\|_1 \le \|w - w_{\mathcal{P}}\|_1 + \|w_{\mathcal{P}} - \mathcal{H}(w_{\mathcal{P}},n)\|_1 + \|\mathcal{H}(w_{\mathcal{P}},n) - H(w,n)\|_1$$

is at most $2\|w - w_{\mathcal{P}}\|_1 \le \epsilon$ as claimed.

For part *ii.*), we rely heavily on [12, Lemma 7.3], which states the following: Let $p_n > 0$ with $p_n \to 0$, $np_n \to \infty$, and for each n let H_n be a weighted graph with n

vertices, unit vertex weights, and contain no loops. Supposing that $\|H_n\|_1$ is uniformly bounded and the edge weights $\beta_{ij}(H)$ satisfy

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{1\le i<j\le n} \max(|\beta_{ij}(H_n)| - p_n^{-1}, 0) = 0, \tag{2.9}$$

then

$$\lim_{n\to\infty} \|(p_n^{-1}\mathsf{G}(H_n, p_n) - H_n\| \quad = 0.$$

We apply this lemma with $H_n = \mathsf{H}(w, n)$. By the theorem of Hoeffding in the proof of $i.$), $\|H_n\|_1 \to \|w\|_1$ almost surely, which implies that $\|H_n\|_1$ is uniformly bounded. Thus we need only check (2.9). We have that

$$\frac{1}{n^2} \sum_{1\le i<j\le n} \max(|\beta_{ij}(H_n)| - p_n^{-1}, 0) = \frac{1}{n^2} \sum_{1\le i<j\le n} \max(|w(x_i, x_j)| - p_n^{-1}, 0)$$

which converges to 0 as $n \to \infty$ with probability 1 by the same theorem of Hoeffding. Thus, as $p_n \to 0$, for every $P > 0$ the limit superior of the above expression is bounded by $\frac{1}{2}\|\max(|w| - P, 0)\|_1$ and this can be mad arbitrarily small by choosing P to be large. Thus, the lemma applies and part $ii.$) of the theorem holds true. □

These w-random graph models provide a great tool for analysis of large networks sampled from graphons. This is because if any sequence of graphs converges to an L^p graphon, that sequence must be close to a w-random sequence of graphs in cut norm.

Proposition 2.3.12 (Proposition 2.16, [12]). *Let* $p > 1$ *and* $\{G_n\}$ *be a sequence of simple graphs such that* $\|G_n\|_1 \to 0$ *and* $\delta \ (G_n/\|G_n\|_1, w) \to 0$*, where* w *is an* L^p *graphon. Let* $G_n' = \mathsf{G}(|V(G_n), w, \|G_n\|_1)$. *Then, with probability 1, one can order the vertices of* G_n *and* G_n' *such that*

$$\left\| \frac{G_n}{\|G_n\|_1} - \frac{G_n'}{\|G_n'\|_1} \right\| \quad \to 0.$$

Proof. By Proposition 2.3.10, the sequence $\{G_n\}$ must be $\|w\|_p$-upper L^p regular. From $\delta\ (G_n/\|G_n\|_1, w) \to 0$ we get that $\|w\|_1 = 1$ and it is shown in [12, Proposition A.1] that L^p upper regular sequences must have unbounded average degree, implying that $n\|G_n\|_1 \to \infty$. It then follows that $\delta\ (G'_n/\|G'_n\|_1, w) \to 0$ with probability 1. It is possible, by [12, Proposition 5.2] to order the vertices of G_n and G'_n so that $d\ (G_n/\|G_n\|_1, w) \to 0$ and $d\ (G'_n/\|G'_n\|_1, w) \to 0$, and thus

$$\left\| \frac{G_n}{\|G_n\|_1} - \frac{G'_n}{\|G'_n\|_1} \right\| \to 0,$$

as desired. □

2.3.5 Counting lemma

While we have discussed multiple results about convergence in δ , a key part of dense graph theory has so far not been shown to have a sparse counterpart—homomorphism densities. Borgs et. al. handle this by developing a ***counting lemma*** for L^p graphons, a claim that any two graphs/graphons close in cut norm must also have close F-densities. However, for general sparse graphs, this is too much to ask for; for example, unlike in the dense case, it is not guaranteed that $t(F, w)$ will be finite. The following proposition shows the conditions for $t(F, w)$ to be finite.

Proposition 2.3.13 (Proposition 2.19, [12])**.** *Let F be a simple graph with maximum degree Δ. For every $p < \Delta$, there exists an L^p graphon w such that $t(F, w) = \infty$. On the other hand, if w is an L^Δ graphon, then $t(F, w)$ is well de ned, nite, and $|t(F, w)| \leq \|w\|_\Delta^{|E(F)|}$.*

With this in mind, any counting lemma will have to account for the maximum degree of F and so no general version can hold. We present such a counting lemma below.

Theorem 2.3.14 (Theorem 2.20, [12]). *Let F be a simple graph with m edges and maximum degree Δ. Let $\Delta < p < \infty$. If u and w are graphons with $\|u\|_p, \|w\|_p \leq 1$ and $\delta \ (u, w) \leq \epsilon$, then*

$$|t(F, u) - t(F, w)| \leq 2m(m - 1 + p - \Delta)\left(\frac{2\epsilon}{p - \Delta}\right)^{\frac{p-\Delta}{p-\Delta+m-1}}.$$

Proof. Let $V(F) = \{1, 2, \ldots, n\}$ and the $E(F) = \{e_1, \ldots, e_m\}$. Let i_t, j_t be the endpoints of e_t for $1 \leq t \leq m$. We assume that $\|u - w\| \ \leq \epsilon$ and then have that

$$\begin{aligned} t(F, u) - t(F, w) &= \int_{[0,1]^n} \left(\prod_{t=1}^{m} u(x_{i_t}, x_{j_t}) - \prod_{t=1}^{m} w(x_{i_t}, x_{j_t})\right) dx_1 \ldots dx_n \\ &= \sum_{t=1}^{m} \int_{[0,1]^n} \left(\prod_{s<t} u(x_{i_s}, x_{j_s})\right)\left(\prod_{s>t} w(x_{i_s}, x_{j_s})\right) u(x_{i_t}, x_{j_t}) dx_1 \ldots dx_n \\ &- \sum_{t=1}^{m} \int_{[0,1]^n} \left(\prod_{s<t} u(x_{i_s}, x_{j_s})\right)\left(\prod_{s>t} w(x_{i_s}, x_{j_s})\right) w(x_{i_t}, x_{j_t}) dx_1 \ldots dx_n. \end{aligned}$$

It then suffices to show that for each $t = 1, \ldots, m$,

$$\begin{aligned} &\left| \int_{[0,1]^n} \left(\prod_{s<t} u(x_{i_s}, x_{j_s})\right)\left(\prod_{s>t} w(x_{i_s}, x_{j_s})\right)(u(x_{i_t}, x_{j_t}) - w(x_{i_t}, x_{j_t})) dx_1, \ldots, dx_n \right| \\ &\leq 2(m - 1 + p - \Delta)\left(\frac{2\epsilon}{p - \Delta}\right)^{\frac{p-\Delta}{p-\Delta+m-1}}. \end{aligned} \tag{2.10}$$

Let $K > 0$, which we will choose later. Let $u = u_{\leq K} + u_{>K}$, where $u_{\leq K} := u\mathbb{1}_{|u| \leq K}$ and $u_{>K} := u\mathbb{1}_{|u|>K}$. Similar definitions hold for $w = w_{\leq K} + w_{>K}$. We claim that

$$\begin{aligned} &\left| \int_{[0,1]^n} \left(\prod_{s<t} u_{\leq K}(x_{i_s}, x_{j_s})\right)\left(\prod_{s>t} w_{\leq K}(x_{i_s}, x_{j_s})\right)(u(x_{i_t}, x_{j_t}) - w(x_{i_t}, x_{j_t})) dx_1, \ldots, dx_n \right| \\ &\leq 4K^{m-1}\epsilon. \end{aligned} \tag{2.11}$$

To show this, we consider the value of x_i to be be fixed for all $i \in [n] \setminus \{i_t, j_t\}$, changing the above integral to the form

$$K^{m-1} \int_{[0,1]^2} (u(x_{i_t}, x_{j_t}) - w(x_{i_t}, x_{j_t})) a(x_{i_t}) b(x_{j_t}) dx_{i_t} dx_{j_t} \tag{2.12}$$

for some functions $a(\cdot)$ and $b(\cdot)$ with $\|a\|_\infty$, $\|b\|_\infty \leq 1$, where $a(\cdot)$ and $b(\cdot)$ depend on the values of x_i for $i \in [n] \setminus \{i_t, j_t\}$ that were fixed. Therefore (2.12) is bounded in absolute value by $K^{m-1}\|u - w\|_{\infty \to 1} \leq 4K^{m-1}\epsilon$ due to the relation between $\| \cdot \|_{\infty \to 1}$ and $\| \cdot \|$. The inequality (2.11) then follows.

Next we claim that the difference between the integrals in (2.10) and (2.11) is bounded in absolute value by $2(m-1)/K^{p-\Delta}$. If we write this difference as a telescoping sum, it suffices to show that each expression of the following form is bounded in absolute value by $2/K^{p-\Delta}$:

$$\int_{[0,1]^n} \left(\prod_{s<t} u_*(x_{i_s}, x_{j_s}) \right) \left(\prod_{s>t} w_*(x_{i_s}, x_{j_s}) \right) (u(x_{i_t}, x_{j_t}) - w(x_{i_t}, x_{j_t})) dx_1, \ldots, dx_n. \tag{2.13}$$

Here we replace exactly one of the $m-1$ subscript $*$s by $> K$, replace some of the other $*$s by $\leq K$, and erase the others. We then consider the special edge e_0 corresponding to the factor whose subscript is replaced by $> K$ and utilize [12, Lemma 8.2] to show that the expression (2.13) is bounded above by

$$\|u_{>K}\|_q \left(\prod_{(s,t) \in \mathcal{I}} \|(u - w)\|_p \right). \tag{2.14}$$

Using the fact that $\|u_{\leq K}\|_p \leq \|u\|_p \leq 1$ and $\|w_{\leq K}\|_p \leq \|w\|_p \leq 1$, the triangle inequality can be used to show that $\|u - w\|_p \leq 2$. Also,

$$\|u_{>K}\|_q \leq \|u(|u|/K)^{p/q-1}\|_q = \|u\|_p^{p/q} / K^{p/q-1} \leq 1/K^{p-\Delta},$$

showing that an integral of the form in (2.14) is bounded above by $2/K^{p-\Delta}$ in absolute value. Combining these bounds, we can show the integral in (2.10) is bounded in absolute value by

$$4K^{m-1}\epsilon + 2(m-1)/K^{p-\Delta}.$$

This bound is optimized with the choice of $K = ((p-\Delta)/(2\epsilon))^{1/(m-1+p-\Delta)}$, which gives the original claimed bound. □

This counting lemma can be used to show convergence of homomorphism densities for sequences of graphons that are uniformly bounded in L^p norm for which L^p upper regularity is not strong enough.

Corollary 2.3.15 (Corollary 2.21, [12])**.** *Let $p > 1$ and $C > 0$, and let $\{w_n\}$ be a sequence of graphons converging to w in cut metric. If $\|w_n\|_p, \|w\|_p \leq C$, then for every simple graph F with maximum degree less than p, we have $t(F, w_n) \to t(F, w)$.*

2.4 Unit square measures as graph limits

Graph limit theory contains two extremes: that of dense graph sequences, and that of very sparse graph sequences, whose maximum degree is universally bounded above. Both extremes have well developed limit theories, with dense graph limit theory introduced in [67] and with very sparse graph limit theory first handled in [11, 45]. It is shown in [55] that the limit objects of these two limit theories can be viewed as symmetric Borel measures on $[0,1]^2$, bringing together the two polar opposites of the graph limits world. Additionally, this newly developed theory is a generalization of L^p sparse limit theory, as any convergent sequence in the latter theory will converge in the former, though the converse is not true. We present some of the results in this paper here to highlight future directions that work in this thesis could take, beginning with some fundamental definitions.

Definition 2.4.1 (Balanced partitions)**.** Let Ω be a finite set of n elements. We say that $\Omega = \bigcup_{i=1}^k \Omega_i$ is a ***balanced partition*** if $|\Omega_i| \in \{\lfloor n/k \rfloor, \lceil n/k \rceil\}$ for $1 \leq i \leq k$.

Every finite set has a balanced partition for every k and the multiset $\{|\Omega_1|, \dots, |\Omega_k|\}$ is uniquely determined by n and k. Furthermore, every partition of Ω has a characteristic $k \times |\Omega|$ matrix M whose rows are the characteristic vectors of the partition sets Ω_i. The sum of each column of M is 1, and if n divides k and the partition is balanced, then each row sum is n/k. We denote the set of all balanced $k \times n$ partition characteristic matrices by $\widetilde{\mathcal{K}}(k, n)$.

Consider an arbitrary $n \times n$ nonnegative matrix S whose rows and columns are indexed by a finite set Ω with n elements with k a fixed natural number. If $\mathcal{P} = \{\Omega_1, \ldots, \Omega_k\}$ is a balanced partition of Ω, we define a matrix $\mathcal{P}(S)$ whose ij-th entry is

$$\sum_{(x,y) \in \Omega_i \times \Omega_j} S_{x,y}.$$

Let $\widetilde{C}(S, k)$ denote the set of all matrices that can be obtained as $\mathcal{P}(S)$ for some balanced partition $\mathcal{P}$ with k sets. We now introduce the idea of the ***shape*** of a matrix, a fundamental concept in this limit theory.

Definition 2.4.2 (Shape of a matrix)**.** Let $\mathcal{K}(k, n)$ denote the space of all nonnegative $k \times n$ matrices with each column sum equal to 1 and each row sum equal to n/k. For an arbitrary matrix S and a fixed natural number k we define the ***shape*** $C(S, k)$ in $\mathsf{R}^{k \times k}$ by

$$C(S, k) := \{MSM^T \mid M \in \mathcal{K}(k, n)\}.$$

We note that as $\mathcal{K}(k, n)$ is compact, $C(S, k)$ is itself a compact subset of the set of real $k \times k$ matrices. $C(S, k)$ is also invariant under conjugation with permutation matrices. For an arbitrary set of matrices $\mathcal{S}$, we define

$$C(\mathcal{S}, k) := \bigcup_{S \in \mathcal{S}} C(S, k).$$

Note that $\widetilde{C}(S, k)$ is not the same space as $C(S, k)$; for example, if S is the identity matrix of size 2, then $\widetilde{C}(S, k)$ is a space of matrices with only integer entries, while $C(S, k)$ has no such restriction. We also note that if $\mathcal{S}$ is a compact set of matrices, then $C(\mathcal{S}, k)$ is also a compact set that depends continuously on $\mathcal{S}$.

It is not too difficult to define an analogue of $C(S, k)$ for a Borel measure μ

on $[0,1]^2$ with finite total measure. Let $f_1, \ldots, f_k$ be nonnegative Borel measurable functions on $[0,1]$ such that their sum is the constant function 1 and for $1 \leq i \leq k$ we have that

$$\int_{[0,1]} f_i \, dx = \frac{1}{k}.$$

For such a sequence of functions, we define a matrix M such that

$$M_{i,j} := \iint_{[0,1]^2} f_i(x) f_j(y) d\mu,$$

sometimes denoting this matrix by $\mathcal{M}(f_1, \ldots, f_k)$. Let $C_0(\mu, k)$ be the set of all $\mathcal{M}(f_1, \ldots, f_k)$ for all possible choices of suitable $f_1, \ldots, f_k$; we denote by $C(\mu, k)$ the topological closure of $C_0(\mu, k)$. It is now finally time to present the limit theory of [55], relying on this background to state their results. Before we begin, as the sum of all entries in a matrix appears often in the following text, we use the shorthand $\gamma(S)$ to represent this quantity, noting that $\gamma(S) = \gamma(X)$ for all $X \in C(S, k)$.

Definition 2.4.3 (Convergence of shapes)**.** Let $c > 0$ be a constant and let $\{S_i\}_{i \geq 1}$ be a sequence of nonnegative matrices that satisfy $\gamma(S_i) < c$ for all i. We say that the sequence $\{S_i\}_{i \geq 1}$ is ***convergent*** if for every natural number k, the shapes $C(S_i, k)$ converge to some fixed closed set in $[0, c]^{k \times k}$ in the Hausdorff topology.

In this context, the Hausdorff distance of compact sets in $\mathsf{R}^{k \times k}$ refers to the metric induced by the ℓ^1 norm of $\mathsf{R}^{k \times k}$. This definition of convergence allows the formulation of the main limit result.

Theorem 2.4.4 (Corollary 4.3, [55])**.** *If $\{S_i\}_{i \geq 1}$ is a convergent sequence of nonnegative symmetric matrices, then there exists a symmetric Borel measure on $[0,1]^2$ such that for every natural number k, the limit shape of $C(S_i, k)$ is $C(\mu, k)$.*

To bring this back to the language of graph limits, we must now analyze how this notion of convergence interacts with graphs themselves. It is thus important to define what the shape of a graph is; we do so below, as well as introduce another important concept related to convergence in this limit theory.

Definition 2.4.5 (s-convergence)**.** For a finite graph G with a non-empty edge set, we define $C_0(G, k)$ to be $\|A_G\|_1^{-1}C_0(A_G, k)$, where A_G is the adjacency matrix of G. With this in mind, we say a sequence of graphs $\{G_n\}_{n\geq 1}$ is ***s-convergent*** if for every fixed natural number k, the shapes $C_0(G_n, k)$ are converging in the Hausdorff metric.

This naturally leads to a direct corollary of Theorem 2.4.4 in the following form.

Corollary 2.4.6 (Theorem 4.5,[55])**.** *If $\{G_n\}_{n\geq 1}$ is an s-convergent sequence of graphs, then there exists a symmetric Borel probability measure μ on $[0, 1]^2$ such that the limit shape of $C_0(G_n, k)$ is $G(\mu, k)$ for all natural numbers k.*

Furthermore, every probability measure of this form arises from a convergent sequence of graphs.

Theorem 2.4.7 (Theorem 4.7, [55])**.** *Let μ be a symmetric Borel probability measure on $[0, 1]^2$. Then there exists a sequence of graphs $\{G_n\}$ such that the limit shape of $C_0(G_n, k)$ is $C(\mu, k)$ for all natural numbers k.*

We show now that this notion of convergence generalizes L^p convergence (and therefore generalizes dense and very sparse convergence as well). Thus, immediate direction for future work along the lines of this thesis would be to develop a graph parameter that identifies Robinson matrices that respects shape convergence (in both directions). Before we present the definition of shape upper L^p regularity, we will need to define several sets and quantities for ease of writing.

We shall let $\mathcal{M}_c^+$ be the set of symmetric Borel measures μ on $[0, 1]^2$ such that $\mu([0, 1]^2) \leq c$. Given a positive integer k, we also define the set $E_k := \{\alpha \in [0, 1]^k \mid \sum_{i=1}^k \alpha_i = 1\}$. Finally, we define a collection of functions in the following way. For a positive integer k, let

$$F_k := \left\{ f \in L^\infty([0, 1])^k \;\middle|\; \sum_{i=1}^k f_i(x) = 1 \;\forall x \in [0, 1] \right\}.$$

Then, for a fixed positive real $\alpha \in \mathsf{R}^k$, we define

$$\mathcal{F}_\alpha := \left\{ f \in F_k \;\middle|\; \int_{[0,1]} f_j = \alpha_j \text{ and } f_j \geq 0 \text{ for all } 1 \leq j \leq k \right\}.$$

With these definitions in mind, we can finally state the equivalence of shape convergence and sparse L^p convergence.

Definition 2.4.8 (Shape upper L^p regularity)**.** Given a positive integer k, an $\alpha \in (0,1]^k$, and a matrix $M \in \mathsf{R}^{k\times k}$, we define the following quantity:

$$\|M\|_{\alpha,p} := \left(\sum_{1\leq i,j\leq k} \left(\frac{M_{i,j}}{\alpha_i\alpha_j}\right)^p \alpha_i\alpha_j \right)^{\frac{1}{p}}.$$

Note that for a family of 0-1 valued functions $f \in \mathcal{F}_\alpha$ defining the partition $\mathcal{P}$ and the graphon w, we have that

$$\|\mathcal{M}(\mu_w, f)\|_{\alpha,p} = \|w_{\mathcal{P}}\|_p,$$

where

$$\mathcal{M}(\mu_w, f)_{i,j} = \iint_{[0,1]^2} f_i(x) f_i(y) w(x,y)\,dxdy.$$

A measure $\mu \in \mathcal{M}_c^+$ is called ***shape*** (C,η)-***upper*** L^p ***regular*** if for any positive integer k, $\alpha \in [\eta,1]^k \cap E_k$ and any family of functions $f \in \mathcal{F}_\alpha$, we have that $\|M(\mu, f)\|_{\alpha,p} \leq C$. A sequence of measures $\{\mu_n\}_{n\geq 1} \subset \mathcal{M}_k^+$ of measures is called ***shape*** C-***upper*** L^p ***regular*** if for any $\eta > 0$ there exists some $n_0 \in \mathsf{N}$ such that for any $n \geq n_0$ the measure μ_n is shape $(C+\eta,\eta)$-upper L^p regular.

Shape upper L^p regularity is clearly stronger than the standard L^p upper regularity as there is no restriction on the family of functions f. However, for graphons it turns out these notions are equivalent and L^p upper regularity implies shape upper L^p regularity. This allows for a proof of the equivalence of convergence for sequences of L^p graphons, showing that this new method of graph limit theory truly encompasses the field.

Theorem 2.4.9 (Corollary 10.9, [55]). *A sequence $\{\mu_n\}_{n\geq 1}$ of absolutely continuous measures in $\mathcal{M}^+$ is C-upper L^p regular if and only if it is shape C-upper L^p regular.*

Chapter 3

TRIANGULAR TRUNCATION: SYMMETRIC MATRICES AND GRAPHONS

3.1 Background

In this chapter, we detail results about the operator $\mathcal{T}_n$, colloquially called the triangular cut or triangular truncation, borrowing heavily from the author's previous work [74]. $\mathcal{T}_n$ takes an $n \times n$ real-valued matrix as its argument and sets all values of said matrix below the main diagonal to 0. The norm growth of this operator is a well-studied problem in operator theory; see for example [2, 89] for explicit calculations and bounds for $\mathcal{T}_n$ applied to matrices equipped with operator norm. $\mathcal{T}_n$ can be viewed as a particular example of a Schur/Hadamard multiplier—a linear operator that multiplies matrices entry-wise with some fixed matrix—and tools available from Schur multiplier theory are used to prove several interesting results about $\mathcal{T}_n$.

Schur multipliers comprise a rich field of study; see [52] for an in-depth look into the field. We refer to [7, 20, 23] for results on the norms of Schur multipliers in general. The triangular cut operator in particular often shows up in various areas of mathematics, from error estimates of certain linear solving methods [89] to the study of group actions on Von-Neumann algebras [26]. It is thus of natural interest to study the behavior of its norm and its interactions with different families of matrices equipped with different matrix norms. It is well known that the operator norm of $\mathcal{T}_n$ grows to infinity when it is considered as an operator on real matrices equipped with the standard operator norm $\|\cdot\|_{\text{opr}}$ (see [2, Theorem 1] for a proof and [88] for estimations of growth speed). In [8], Bennett showed that when $1 < p < q < \infty$, the triangular cut acting on linear maps from $\ell^p(\mathsf{R})$ to $\ell^q(\mathsf{R})$ is bounded. Recently, Coine used the canonical characterization of Schur multipliers to prove that the triangular cut acting on linear maps from $\ell^p(\mathsf{R})$ to $\ell^q(\mathsf{R})$ is unbounded when $1 \leq q \leq p \leq \infty$ [20].

A generalized version of the triangular cut, denoted M_χ, can be defined on the space of real-valued functions on $[0,1]^2$ in the following way: Let w be a function mapping $[0,1]^2$ to the reals and let M_χ map $w(x,y)$ to $w(x,y)\chi(x,y)$, where $\chi(x,y) = 0$ if $x < y$ and 1 otherwise. M_χ is a multiplication operator, a type of linear operator that multiples functions by another fixed function, in this case a characteristic function of a set in R^2.

Similar to the case of matrices, we prove here that the norm of the triangular cut on functions with respect to the cut norm is infinite, in particular showing that M_χ is unbounded even when restricted to graphons, symmetric measurable functions acting on $[0,1]^2$. To make conclusive statements about the behavior of M_χ on graphons, we must know how the norm of $\mathcal{T}_n$ grows as an operator on symmetric matrices. We could not find clear reference in existing literature for any such analysis, with respect to the cut norm or otherwise, and so we address this issue here. To do so, we note that the cut norm is equivalent to an injective tensor norm. This allows us to make use of some bounds and techniques from [7] to show that the norm of $\mathcal{T}_n$ on symmetric matrices grows to infinity.

We make a special note that the generalization of $\mathcal{T}_n$ to M_χ is inspired by the natural identification between matrices and step functions in $L^\infty([0,1]^2)$. In fact, the completion of the space of matrices under the cut norm becomes the space $L^\infty([0,1]^2)$. This natural relationship between these two spaces allows for results proved about $\mathcal{T}_n$ to be extended to results about M_χ. In particular, as symmetric matrices form a dense subspace in the space of graphons, analyzing how $\mathcal{T}_n$ acts on this subspace will provide insight about how M_χ will act on graphons.

For the sake of completeness, we also present a proof of the unboundedness of $\mathcal{T}_n$ on symmetric matrices with respect to the standard operator norm. Furthermore, we introduce a new operator related to the triangular cut that we call the ***banded*** cut, and we prove it is an unbounded operator when acting on graphons. Similar techniques used for the proofs of how $\mathcal{T}_n$ behaves on matrices can be used to state analogous results for the banded cut and we omit these proofs from the text.

3.2 Notation

In this subsection we collect the necessary background for the chapter. We denote the space of $n \times n$ real-valued matrices as $M_n(\mathsf{R})$ and the space of symmetric $n \times n$ real-valued matrices as $\mathcal{S}_n(\mathsf{R})$. For vectors x in R^n and functions f from $[0,1]^2$ to R, we denote the standard p-norm with $\|\cdot\|_p$, respectively defined as

$$\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$
$$\|f\|_p := \left(\int_{[0,1]^2} |f|^p dxdy\right)^{1/p}.$$

The set $\{1, ..., n\}$ shall be denoted $[n]$ and tensor products of matrices shall be denoted by the symbol $\otimes$. As a reminder, if $A = [a_{ij}]$ is an $n \times m$ matrix and B is a $p \times q$ matrix, then their tensor product is the $np \times mq$ block matrix $A \otimes B$, where

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1m}B \\ a_{21}B & a_{22}B & \dots & a_{2m}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}B & a_{n2}B & \dots & a_{nm}B \end{bmatrix}.$$

We note that the tensor product is not commutative, a property of much consternation for the author. The main matrix norms used throughout the chapter are defined next. Let A be an $n \times n$ matrix. We denote the ***cut norm*** of A, introduced in [38], by

$$\|A\|\quad = \frac{1}{n^2} \max_{A,B \subseteq [n]} \left| \sum_{i \in A,\ j \in B} a_{ij} \right|,$$

and the (p,q)-***norm*** of A [7, Equation 1] by

$$\|A\|_{\mathcal{B}(\ell^p,\ell^q)} = \sup_{\|x\|_p \le 1} \|Ax\|_q.$$

We use the more conventional notation $\|A\|_{\text{opr}}$ to denote the $(2,2)$-norm, which is simply called the ***operator norm*** (sometimes the ***spectral norm***) of A. Lastly, we introduce the cast of operators that feature in the chapter, starting by letting $T_n \in M_n(\mathsf{R})$ be

$$(T_n)_{ij} = \begin{cases} 1 & i \le j \\ 0 & i > j \end{cases}, \tag{3.1}$$

and defining the ***triangular cut for matrices*** as

$$\mathcal{T}_n : M_n(\mathsf{R}) \to M_n(\mathsf{R}), \quad A \mapsto A \circ T_n, \tag{3.2}$$

where $\circ$ represents Schur multiplication (i.e. entrywise multiplication). To generalize $\mathcal{T}_n$ to graphons, we let $\chi : [0,1]^2 \to [0,1]$ be defined as

$$\chi(x,y) = \begin{cases} 1 & x \le y \\ 0 & x > y \end{cases}, \tag{3.3}$$

and define the ***triangular cut on graphons*** as

$$M_\chi : L^p([0,1]^2) \to L^p([0,1]^2), \quad w \mapsto w\chi. \tag{3.4}$$

3.3 Norm estimates

In this subsection, we show that the norm of the linear operator $\mathcal{T}_n$ defined by the triangular cut on matrices equipped with cut norm grows to infinity. Prefacing our result, we need the following proposition, which may be considered folklore, but we present a proof for the sake of completeness. The proof of part (2) can be found in [66] for the unscaled version of the cut norm. The proof of part (3) was inspired by [8].

Proposition 3.3.1. *For $n \times n$ matrices A, B, we have*

1. $n\|A\| \quad \le \|A\|_{opr}$.

2. $n^2\|A\|_\square \le \|A\|_{\mathcal{B}(\ell^\infty,\ell^1)} \le 4n^2\|A\|_\square$.

3. $\|A\|_{\mathcal{B}(\ell^\infty,\ell^1)}\|B\|_{\mathcal{B}(\ell^\infty,\ell^1)} \le \|A\otimes B\|_{\mathcal{B}(\ell^\infty,\ell^1)} \le \frac{\pi}{2}K_G^2\|A\|_{\mathcal{B}(\ell^\infty,\ell^1)}\|B\|_{\mathcal{B}(\ell^\infty,\ell^1)}$, *where the quantity* $1.676 < K_G < 1.782$ *is the Grothendieck constant.*

Proof. For 1, note the following:

$$\begin{aligned} n^2\|A\|_\square &= \max_{S,T\subseteq[n]}\left|\sum_{i\in S,j\in T} a_{ij}\right| = n\max_{S,T\subseteq[n]}\left|\left\langle A\frac{\mathbb{1}_S}{\sqrt{n}},\frac{\mathbb{1}_T}{\sqrt{n}}\right\rangle\right| \\ &\le n\sup_{x,y\in\mathsf{R}^n}\left|\left\langle A\frac{x}{\|x\|_2},\frac{y}{\|y\|_2}\right\rangle\right| = n\|A\|_{\mathrm{opr}}, \end{aligned}$$

where $\mathbb{1}_S$ is the indicator function of the set S identified with a vector in R^n. To prove 2, first note that by definition,

$$\|A\|_{\mathcal{B}(\ell^\infty,\ell^1)} = \sup_{\|x\|_\infty\le1}\|Ax\|_1 = \sup_{\|x\|_\infty\le1}\sum_{i=1}^n\left|\sum_{j=1}^n a_{ij}x_j\right| = \sup_{\|x\|_\infty,\|y\|_\infty\le1}\left|\sum_{i=1}^n\sum_{j=1}^n a_{ij}x_jy_i\right|,$$

and further refer to [66, Lemma 8.10] for an equivalent definition of the cut norm of A:

$$\|A\|_\square = \frac{1}{n^2}\sup_{0\le x_i,y_j\le1}\left|\sum_{i,j}a_{ij}x_iy_j\right|,$$

showing that the lower bound holds. For the upper bound, we rewrite $\|A\|_{\mathcal{B}(\ell^\infty,\ell^1)}$ in the following way:

$$\begin{aligned} \|A\|_{\mathcal{B}(\ell^\infty,\ell^1)} &= \sup_{0\le x_i,z_i,y_j,w_j\le1}\left|\sum_{i,j}a_{ij}(x_i-z_i)(y_j-w_j)\right| \\ &\le \sup_{0\le x_i,z_i,y_j,w_j\le1}\left(\left|\sum_{i,j}a_{ij}x_iy_j\right| + \left|\sum_{i,j}a_{ij}x_iw_j\right| + \left|\sum_{i,j}a_{ij}z_iy_j\right| + \left|\sum_{i,j}a_{ij}z_iw_j\right|\right) \\ &\le \sup_{\|x\|_\infty,\|y\|_\infty\le1}\left|\sum_{i,j}a_{ij}x_iy_j\right| + \sup_{\|x\|_\infty,\|w\|_\infty\le1}\left|\sum_{i,j}a_{ij}x_iw_j\right| \\ &+ \sup_{\|z\|_\infty,\|y\|_\infty\le1}\left|\sum_{i,j}a_{ij}z_iy_j\right| + \sup_{\|z\|_\infty,\|w\|_\infty\le1}\left|\sum_{i,j}a_{ij}z_iw_j\right| \\ &= 4n^2\|A\|_\square , \end{aligned}$$

proving the original claim. To prove 3, consider the vectors $x, y, z, w \in \mathsf{R}^n$, and note that

$$
\begin{aligned}
|\langle y, Ax\rangle| \cdot |\langle w, Bz\rangle| &= |\langle (y \otimes w), (A \otimes B)(x \otimes z)\rangle| \\
&\leq \|A \otimes B\|_{\mathcal{B}(\ell^\infty, \ell^1)} \|x \otimes z\|_\infty \|y \otimes w\|_\infty \\
&= \|A \otimes B\|_{\mathcal{B}(\ell^\infty, \ell^1)} \|x\|_\infty \|y\|_\infty \|z\|_\infty \|w\|_\infty,
\end{aligned}
$$

from which the lower bound follows. The upper bound was proved in [7, Proposition 10.2], which we summarize here. Using a result of Pietsch [79, Theorem 11], the authors first show

$$
\|A \otimes B\|_{\mathcal{B}(\ell^2, \ell^q)} \leq K(q) \|A\|_{\mathcal{B}(\ell^2, \ell^q)} \|B\|_{\mathcal{B}(\ell^2, \ell^q)},
$$

where $K(q)$ is a constant dependent only on q. From there, the authors improve Pietsch's bound on $K(q)$ by following his arguments using normalized independent Gaussian random variables instead of Rademacher functions to get that $K(q) = \pi^{1/2q} 2^{-1/2} \Gamma(\frac{1}{2}(q+1))^{1/q}$ for real scalars. Taking transposes and using the identity $(A \otimes B)^* = A^* \otimes B^*$, we get that

$$
\|A \otimes B\|_{\mathcal{B}(\ell^p, \ell^2)} \leq K(p^*) \|A\|_{\mathcal{B}(\ell^p, \ell^2)} \|B\|_{\mathcal{B}(\ell^p, \ell^2)},
$$

where p^* is the conjugate dual of p. For the general case, a result of Lindenstrauss and Pelczynski [63, p.321] guarantees for a given A, B the existence of matrices A_1, A_2, B_1, B_2 such that $A = A_1 \circ A_2$, $B = B_1 \circ B_2$, and

$$
\begin{aligned}
\|A_1\|_{\mathcal{B}(\ell^2, \ell^q)} \|A_2\|_{\mathcal{B}(\ell^p, \ell^2)} &\leq K_G \|A\|_{\mathcal{B}(\ell^p, \ell^q)} \\
\|B_1\|_{\mathcal{B}(\ell^2, \ell^q)} \|B_2\|_{\mathcal{B}(\ell^p, \ell^2)} &\leq K_G \|B\|_{\mathcal{B}(\ell^p, \ell^q)}.
\end{aligned}
$$

Thus, using [7, Proposition 2.1], we have the following:

$$
\begin{aligned}
\|A \otimes B\|_{\mathcal{B}(\ell^\infty, \ell^1)} = \|(A_1 \circ A_2) \otimes (B_1 \circ B_2)\|_{\mathcal{B}(\ell^\infty, \ell^1)} &= \|(A_1 \otimes B_1) \circ (A_2 \otimes B_2)\|_{\mathcal{B}(\ell^\infty, \ell^1)} \\
&\leq \|A_1 \otimes B_1\|_{\mathcal{B}(\ell^2, \ell^1)} \|A_2 \otimes B_2\|_{\mathcal{B}(\ell^\infty, \ell^2)}
\end{aligned}
$$

$$\leq \frac{\pi}{2}\|A_1\|_{\mathcal{B}(\ell^2,\ell^1)}\|B_1\|_{\mathcal{B}(\ell^2,\ell^1)}\|A_2\|_{\mathcal{B}(\ell^\infty,\ell^2)}\|B_2\|_{\mathcal{B}(\ell^\infty,\ell^2)}$$
$$\leq \frac{\pi}{2}K_G^2\|A\|_{\mathcal{B}(\ell^\infty,\ell^1)}\|B\|_{\mathcal{B}(\ell^\infty,\ell^1)},$$

proving the upper bound. □

We now state and prove our norm estimates for $\mathcal{T}_n$ on several different spaces of real-valued matrices.

Proposition 3.3.2. *For $\mathcal{T}_n$ as in (3.2), we have:*

1. $\sup_{A\in M_n(\mathsf{R})} \frac{\|\mathcal{T}_n(A)\|}{\|A\|} \to \infty$ *as* $n \to \infty$.
2. $\sup_{A\in \mathcal{S}_n(\mathsf{R})} \frac{\|\mathcal{T}_n(A)\|}{\|A\|} \to \infty$ *as* $n \to \infty$.
3. $\sup_{A\in \mathcal{S}_n(\mathsf{R})} \frac{\|\mathcal{T}_n(A)\|_{\text{opr}}}{\|A\|_{\text{opr}}} \to \infty$ *as* $n \to \infty$.

Proof. Consider the $n \times n$ matrix A_n defined as below:

$$(A_n)_{ij} = \begin{cases} 0 & i = j \\ (i-j)^{-1} & i \neq j \end{cases}. \tag{3.5}$$

First, we calculate $\|A_n\|$. By Proposition 3.3.1 part 1, we have an upper bound given by $n\|A_n\| \leq \|A_n\|_{\text{opr}}$. Thus, we need only bound $\|A_n\|_{\text{opr}}$, which we do similarly to [2] by introducing a function f on the unit circle such that $f(e^{i\theta}) = i(\pi - \theta)$. The Fourier series of this function is

$$\sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{1}{n}e^{in\theta}.$$

Let T_f be the corresponding multiplication operator that acts on H^2, the space of holomorphic functions on the unit disk with finite 2-norm. The matrix for T_f with respect to the basis $\{e^{in\theta}\}_{n=0}^{\infty}$ for H^2, letting the rows and columns be indexed by those basis elements, is

$$T_f = \begin{bmatrix} 0 & -1 & -\frac{1}{2} & -\frac{1}{3} & \dots \\ 1 & 0 & -1 & -\frac{1}{2} & \dots \\ \frac{1}{2} & 1 & 0 & -1 & \dots \\ \frac{1}{3} & \frac{1}{2} & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{3.6}$$

of which A_n is clearly a submatrix. Thus, because $\|A_n\|_{\text{opr}} \leq \|T_f\|_{\text{opr}}$, we can now focus on bounding $\|T_f\|_{\text{opr}}$. A standard theorem of multiplication operators [28] can be used to show that $\|T_f\|_{\text{opr}} = \sup_\theta |f(e^{i\theta})| = \pi$. Thus, $\|A_n\|_{\text{opr}} \leq \pi$, and so it must be that $\|A_n\| \leq \frac{\pi}{n}$. Now we shall calculate $\|\mathcal{T}_n(A_n)\|$ directly; this is relatively simple, as the matrix is strictly negative above the diagonal. We get

$$\|\mathcal{T}_n(A_n)\| = \frac{1}{n^2} \sum_{1\leq i<j\leq n} (A_n)_{ij} = \frac{1}{n^2} \sum_{i=0}^{n-1} \frac{i}{n-i} = \frac{n(H_{n-1}-1)+1}{n^2}, \tag{3.7}$$

letting H_n denote the n-th harmonic number. This statement can be shown using mathematical induction. For $n = 2$ we have $\sum_{i=0}^{1} i/(2-i) = 1 = 2(H_1 - 1) + 1$, and we further note that

$$\begin{aligned} \sum_{i=0}^{n} \frac{i}{n+1-i} = \sum_{i=1}^{n} \frac{i}{n+1-i} = \sum_{i=0}^{n-1} \frac{i+1}{n-i} = \sum_{i=0}^{n-1} \frac{i}{n-i} + H_n \\ = n(H_{n-1}-1) + 1 + H_n = (n+1)H_n - n, \end{aligned}$$

proving the claim. Thus we have

$$\sup_{A\in M_n(\mathsf{R})} \frac{\|\mathcal{T}_n(A)\|}{\|A\|} \geq \frac{\|\mathcal{T}_n(A_n)\|}{\|A_n\|} \geq \frac{n^2(H_{n-1}-1)+n}{n^2} \to \infty,$$

proving (i).

To prove (ii), consider again A_n as defined in (3.5), and note that $A_n \otimes A_n$ is a symmetric matrix. Using Proposition 3.3.1 part 2 and part 3, we have

$$\begin{aligned} n^4\|A_n \otimes A_n\| &\leq \|A_n \otimes A_n\|_{\mathcal{B}(\ell^\infty,\ell^1)} \leq \frac{\pi}{2}K_G^2\|A_n\|_{\mathcal{B}(\ell^\infty,\ell^1)}\|A_n\|_{\mathcal{B}(\ell^\infty,\ell^1)} \\ &\leq 2\pi K_G^2 n^2\|A_n\| \ \|A_n\|_{\mathcal{B}(\ell^\infty,\ell^1)} \leq 2\pi^2 K_G^2 n\|A_n\|_{\mathcal{B}(\ell^\infty,\ell^1)}, \end{aligned}$$

recalling that $\|A_n\| \leq \frac{\pi}{n}$. Before continuing, we must make note of the fact that $T_{n^2} \circ (A_n \otimes A_n) = (T_n \circ A_n) \otimes A_n$. This is not generally true for any matrix but holds true here because A_n is identically zero on the diagonal. For a matrix B where this is not true, there is potential for the triangular cut T_{n^2} to preserve entries in $B \otimes B$ that in $(T_n \circ B) \otimes B$ would vsnish. With this fact in mind, by item 2 of Proposition 3.3.1,

$$\begin{aligned} 4n^4\|T_{n^2} \circ (A_n \otimes A_n)\| &\geq \|T_{n^2} \circ (A_n \otimes A_n)\|_{\mathcal{B}(\ell^\infty,\ell^1)} = \|(T_n \circ A_n) \otimes A_n\|_{\mathcal{B}(\ell^\infty,\ell^1)} \\ &\geq \|T_n \circ A_n\|_{\mathcal{B}(\ell^\infty,\ell^1)}\|A_n\|_{\mathcal{B}(\ell^\infty,\ell^1)} \geq n^2\|T_n \circ A_n\| \ \|A_n\|_{\mathcal{B}(\ell^\infty,\ell^1)} \\ &= (n(H_{n-1} - 1) + 1)\|A_n\|_{\mathcal{B}(\ell^\infty,\ell^1)}, \end{aligned}$$

where the last equality was shown in (3.7). Thus,

$$\sup_{A \in \mathcal{S}_{n^2}} \frac{\|\mathcal{T}_{n^2}(A)\|}{\|A\|} \geq \frac{\|\mathcal{T}_{n^2}(A_n \otimes A_n)\|}{\|A_n \otimes A_n\|} \geq \frac{n(H_{n-1} - 1) + 1}{8\pi^2 K_G^2 n} \to \infty,$$

proving (ii).

To prove (iii), we make use of the well-known fact that for any matrices A, B we have $\|A \otimes B\|_{\text{opr}} = \|A\|_{\text{opr}}\|B\|_{\text{opr}}$, as well as once again considering the symmetric matrix $A_n \otimes A_n$. It is clear that $\|A_n \otimes A_n\|_{\text{opr}} = \|A_n\|_{\text{opr}}\|A_n\|_{\text{opr}} \leq \pi\|A_n\|_{\text{opr}}$, where $\|A_n\|_{\text{opr}} \leq \pi$ was shown in the proof of (i). On the other hand,

$$\|T_{n^2} \circ (A_n \otimes A_n)\|_{\text{opr}} = \|(T_n \circ A_n) \otimes A_n\|_{\text{opr}} = \|T_n \circ A_n\|_{\text{opr}}\|A_n\|_{\text{opr}}.$$

Next, we show $\|T_n \circ A_n\|_{\text{opr}} \geq \frac{4}{5}\log n$ by considering $T_n \circ A_n$ applied to the carefully chosen vector $v = (n-1)^{-1/2}\sum_{k=2}^n e_k$, where $\{e_n\}_n$ form the canonical basis of R^n. We begin by noting that

$$\|(T_n \circ A_n)v\|_2^2 = \sum_{i=1}^{n}[(T_n \circ A_n)v]_i^2 = \sum_{i=1}^{n}\left(\sum_{j=1}^{i}\frac{1}{j\sqrt{n-1}}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{i}\frac{1}{j}\right)^2$$
$$= \left[\sum_{i=1}^{n}\frac{1}{(n-1)^2}\right]\left[\sum_{i=1}^{n}\left(\sum_{j=1}^{i}\frac{1}{j}\right)^2\right].$$

From here, an application of the Cauchy-Schwartz inequality gives

$$\left[\sum_{i=1}^{n}\frac{1}{(n-1)^2}\right]\left[\sum_{i=1}^{n}\left(\sum_{j=1}^{i}\frac{1}{j}\right)^2\right] \geq \sum_{i=1}^{n}\sum_{j=1}^{i}\frac{1}{j(n-1)} = \frac{1}{n-1}\sum_{i=0}^{n}\frac{i}{n-i+1} = \frac{(n+1)H_n - n}{n-1},$$

which is greater than or equal to $4/5\log n$ for all n greater than or equal to 2. Thus we have

$$\sup_{A\in\mathcal{S}_{n^2}}\frac{\|\mathcal{T}_{n^2}(A)\|_{\text{opr}}}{\|A\|_{\text{opr}}} \geq \frac{\|\mathcal{T}_{n^2}(A_n\otimes A_n)\|_{\text{opr}}}{\|A_n\otimes A_n\|_{\text{opr}}} \geq \frac{4}{5\pi}\log n \to \infty,$$

proving (iii). □

The above results are powerful and can be generalized to the function space $L^p([0,1]^2)$ via a natural embedding of the example matrices used in the proof. Of particular importance is 3.3.2 part (ii), as it can be used to provide analysis of the triangular cut's behavior on graphons, a topic of significant interest in graph limit theory.

3.4 The triangular and banded cut on graphons

When the triangular cut is applied to a matrix, it will output an upper-triangular matrix, a subclass of matrices that are of much mathematical interest. Another family of matrices that receive attention is the family of banded matrices: These are matrices A such that there exists a constant $k \in \mathsf{N}$ where $A_{i,j} = 0$ if $|i-j| > k$. Banded matrices are useful in fast computations of numerical approximations, such as the finite-element methods as applied to certain partial differential equations [50, Section 1.9]. With regards to graph limits, transforming a matrix into a banded matrix can be viewed as an operator that deletes certain sets of edges from a graph, as the transformation will delete entry $(A_G)_{ij}$ if and only if it also deletes entry $(A_G)_{ji}$. In this section, we use our results on the triangular cut on matrices to show that both the triangular and *banded* cut on graphons are unbounded. We begin by handling the triangular cut on graphons.

Proposition 3.4.1. *Let M_χ be as de ned in (3.4). Then for all $c > 0$, there exists a graphon $w \in \mathcal{W}^\infty$ such that $\|M_\chi(w)\| > c\|w\|$. In other words, $\|M_\chi\|_{\text{opr}} = \infty$ on the space $(\mathcal{W}^\infty, \|\cdot\|)$. Furthermore, since $(\mathcal{W}^\infty, \|\cdot\|) \subset (\mathcal{W}^p, \|\cdot\|)$ for all $1 \leq p < \infty$, M_χ acting on $(\mathcal{W}^p, \|\cdot\|)$ is also an unbounded operator.*

Proof. Let $c > 0$ be fixed. Then by Proposition 3.3.2 (ii) there exists an $n \in \mathsf{N}$ such that $\|\mathcal{T}_{n^2}(A_n \otimes A_n)\| > c\|A_n \otimes A_n\|$, where A_n is the matrix defined in (3.5). Let $I_i = [\frac{i-1}{n^2}, \frac{i}{n^2})$ for $1 \leq i \leq n^2$ and define the graphon w_n as

$$w_n(x, y) = (A_n \otimes A_n)_{ij} \quad \text{if } (x, y) \in I_i \times I_j \tag{3.8}$$

It is clear that $\|M_\chi(w_n)\| = \|\mathcal{T}_{n^2}(A_n \otimes A_n)\|$ and that $\|w_n\| = \|A_n \otimes A_n\|$. Thus, it must be that $\|M_\chi(w_n)\| > c\|w_n\|$, proving the claim and showing that M_χ is unbounded on $(\mathcal{W}^\infty, \|\cdot\|)$. Moreover, because $w_n \in \mathcal{W}^p$ for all $p \geq 1$, it is clear that M_χ is also unbounded on $(\mathcal{W}^p, \|\cdot\|)$ for all such p. □

Remark 3.4.2. In [19], the authors prove that for graphons w such that $w : [0, 1]^2 \to [-2, 2]$, $\|w\chi\| \leq 2\sqrt{\|w\|}$. However, we must be careful to note that their result does not imply continuity of M_χ on $(\mathcal{W}^\infty, \|\cdot\|)$, as this would imply that the square root was unnecessary. Clearly this cannot happen due to Corollary 3.4.1, and so we make note that this result can only be extended to bounded subsets of $\mathcal{W}^\infty$. This is because following the proof in [19], we get that $\|w\chi\| \leq \|w\|_\infty\sqrt{\|w\|}$, showing that this technique fails for $\mathcal{W}^\infty$ in general.

We now handle the banded cut, beginning with its definition.

Definition 3.4.3 (Banded cut on graphons)**.** Let $0 < \lambda < 1$ and let the characteristic function $\mathbb{1}_\lambda : [0,1]^2 \to [0,1]$ be defined as

$$\mathbb{1}_\lambda(x,y) = \begin{cases} 1 & |x-y| \leq 1-\lambda \\ 0 & \text{otherwise,} \end{cases}$$

and define the banded cut on graphons as

$$B_\lambda : \mathcal{W}^p \to \mathcal{W}^p, \quad w \mapsto w\mathbb{1}_\lambda.$$

One can view this cut as snipping off two triangles of equal size from opposite corners of the unit square (see Figure 3.1). We now introduce a lemma that will prove useful for our proof of the unboundedness of the banded cut.

Lemma 3.4.4. *Let* $w \in L^p([0,1]^2)$ *and for* $0 < \lambda \leq \frac{1}{2}$*, de ne the two functions* $w_\lambda, \overline{w}_\lambda \in L^p([0,1]^2)$ *as*

$$w_\lambda(x,y) = \begin{cases} w(\frac{1}{\lambda}x, \frac{1}{\lambda}y - \frac{1-\lambda}{\lambda}) & (x,y) \in [0,\lambda] \times [1-\lambda, 1] \\ 0 & \textit{else} \end{cases}$$
$$\overline{w}_\lambda(x,y) = w_\lambda(y,x).$$

Then $\|w_\lambda + \overline{w}_\lambda\| \;\; = 2\lambda^2 \|w\| \;\;$*.*

Proof. We first show that $\|w_\lambda\| \;\; = \|\overline{w}_\lambda\| \;\; = \lambda^2\|w\| \;\;$. It is clear that $\|w_\lambda\| \;\; = \|\overline{w}_\lambda\| \;\;$; if $S \times T$ is the set that achieves $\|w_\lambda\| \;\;$, then $T \times S$ is the set that achieves $\|\overline{w}_\lambda\| \;\;$. Thus, we shall only prove that $\|w_\lambda\| \;\; = \lambda^2\|w\| \;\;$. First we show that $\lambda^2\|w\| \;\; \leq \|w_\lambda\| \;\;$. Let $S \times T \subset [0,1]^2$ be the set that achieves $\|w\| \;\;$. Then define the sets S', T' so that $S' = \{\lambda x | x \in S\}$ and $T' = \{\lambda y + 1 - \lambda | y \in T\}$. With this in mind, it is clear that

$$\lambda^2\|w\| \;\; = \lambda^2 \left| \iint_{S\times T} w \; dxdy \right| = \left| \iint_{S'\times T'} w_\lambda dxdy \right| \leq \|w_\lambda\| \;\;,$$

proving the upper bound. Now we consider the lower bound by letting $U \times V \subset [0,1]^2$ be the set which achieves $\|w_\lambda\| \;\;$. We proceed similarly, defining sets $U' = \{x/\lambda | x \in U\}$

and $V\{y/\lambda - (1-\lambda)\lambda | y \in V\}$. This allows us to show that

$$\|w_\lambda\| \ = \left| \iint_{U \times V} w_\lambda dxdy \right| = \lambda^2 \left| \iint_{U' \times V'} w \ dxdy \right| \leq \lambda^2 \|w\| \ ,$$

proving the lower bound and showing equality of $\lambda^2\|w\|$ and $\|w_\lambda\|$. The rest of the proof proceeds in a similar fashion, starting with the following upper bound argument:

$$\|w_\lambda + \overline{w}_\lambda\| \ \leq \|w_\lambda\| \ + \|\overline{w}_\lambda\| \ = \lambda^2\|w\| \ + \lambda^2\|w\| \ = 2\lambda^2\|w\| \ .$$

For a lower bound, let $S_1 \times T_1 \subseteq [0,\lambda] \times [1-\lambda, 1]$ and let $S_2 \times T_2 \subseteq [1-\lambda, 1] \times [0, \lambda]$. We make note of the fact that

$$\begin{aligned} \|w_\lambda + \overline{w}_\lambda\| \ &\geq \left| \iint_{(S_1 \cup S_2) \times (T_1 \cup T_2)} w_\lambda + \overline{w}_\lambda dxdy \right| \\ &= \left| \iint_{S_1 \times T_1} (w_\lambda + \overline{w}_\lambda) dxdy + \iint_{S_1 \times T_2} (w_\lambda + \overline{w}_\lambda) dxdy \right. \\ &+ \left. \iint_{S_2 \times T_1} (w_\lambda + \overline{w}_\lambda) dxdy + \iint_{S_2 \times T_2} w_\lambda + \overline{w}_\lambda \right| \\ &= \left| \iint_{S_1 \times T_1} (w_\lambda + \overline{w}_\lambda) dxdy + \iint_{S_2 \times T_2} (w_\lambda + \overline{w}_\lambda) dxdy \right| \\ &= \left| \iint_{S_1 \times T_1} w_\lambda dxdy + \iint_{S_2 \times T_2} \overline{w}_\lambda dxdy \right|, \end{aligned} \tag{3.9}$$

as $w_\lambda + \overline{w}_\lambda = 0$ on both $S_1 \times T_2$ and $S_2 \times T_1$, $w_\lambda = 0$ on $S_2 \times T_2$, and $\overline{w}_\lambda = 0$ on $S_1 \times T_1$. Letting $S_1 \times T_1$ and $S_2 \times T_2$ be the sets on which $\|w_\lambda\|$ and $\|\overline{w}_\lambda\|$ are respectively achieved, and noting that $\|w_\lambda\| \ = \|\overline{w}_\lambda\| \ = \lambda^2\|w\|$, we can use (3.9) to show

$$\|w_\lambda + \overline{w}_\lambda\| \ \geq \|w_\lambda\| \ + \|\overline{w}_\lambda\| \ = 2\lambda^2\|w\| \ ,$$

proving the original claim. □

This lemma allows us to make use of a specific auxiliary function that will be used to prove the unboundedness of the banded cut, as shown in the following corollary.

Corollary 3.4.5. *Let $1 \leq p \leq \infty$. Then for $0 < \lambda \leq \frac{1}{2}$, B_λ is an unbounded operator on $(\mathcal{W}^p, \|\cdot\|\ \ )$. In other words, for any $c > 0$, there exists a graphon $w \in \mathcal{W}^p$ such that $\|B_\lambda(w)\|\ \ > c\|w\|\ \ $.*

The idea of the following proof is to make use of the fact that the triangular cut is unbounded on graphons by forcing the banded cut to act like the triangular cut on two graphons at once. For a visual explanation of how this is done, refer to Figure 3.1.

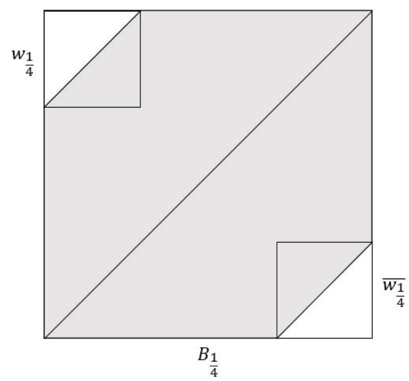

Figure 3.1: An example of the banded cut: the operator $B_{\frac{1}{4}}$ acting on the function $w_{\frac{1}{4}} + \overline{w}_{\frac{1}{4}}$. Gray regions are unchanged while white regions are set to zero.

Proof. Let $0 < \lambda \leq \frac{1}{2}$ be fixed, let $n \geq 2$, let w_n be the graphon defined in (3.8), and let A_n be the matrix defined in (3.5). We will now consider the behavior of B_λ applied to the graphon $(w_n)_\lambda + (\overline{w_n})_\lambda$, making liberal use of Lemma 3.4.4. We first note that

$$\|(w_n)_\lambda + (\overline{w_n})_\lambda\|\ \ = 2\lambda^2\|w_n\|\ \ = 2\lambda^2\|A_n \otimes A_n\|\ \ \leq \frac{4\lambda^2\pi^2K_G^2}{n^3}\|A_n\|_{\mathcal{B}(\ell^\infty,\ell^1)},$$

where the last inequality was shown in Proposition 3.3.2. Furthermore, we also note that $B_\lambda((w_n)_\lambda + (\overline{w_n})_\lambda) = (w_n\chi)_\lambda + (\overline{w_n\chi})_\lambda$. So

$$\|B_\lambda((w_n)_\lambda + (\overline{w_n})_\lambda)\|\ \ = \|(w_n\chi)_\lambda + (\overline{w_n\chi})_\lambda\|\ \ = 2\lambda^2\|w_n\chi\|\ \ ,$$

which is equal to $2\lambda^2\|A_n \otimes A_n\|$ by construction of w_n. Thus,

$$2\lambda^2\|A_n \otimes A_n\| \geq \frac{\lambda^2(n(H_{n-1}-1)+1)}{2n^4}\|A_n\|_{\mathcal{B}(\ell^\infty,\ell^1)},$$

where the last inequality was also shown in Proposition 3.3.2, implies that

$$\sup_{w\in\mathcal{W}^p}\frac{\|B_\lambda(w)\|}{\|w\|} \geq \frac{\|B_\lambda(w_n)\|}{\|w_n\|} \geq \frac{n(H_{n-1}-1)+1}{8\pi^2K_G^2 n} \to \infty.$$

This shows that for $1 \leq p \leq \infty$ and $0 < \lambda \leq \frac{1}{2}$, $\|B_\lambda\|_{\text{opr}} = \infty$ on $(\mathcal{W}^p, \|\cdot\|)$, proving the corollary. □

In terms of graphs, one can think of the operator B_λ as deleting a proportion of the edges of a graph equal to λ^2. More specifically, B_λ will delete edges between ordered vertices whose labels are greater than $n\lambda$ apart from each other. Corollary 3.4.5 shows that this operation will not necessarily respect graph limits for any proportion $0 < \lambda \leq 1/2$ of edges deleted in this manner.

Chapter 4

MEASURING THE ROBINSON PROPERTY

4.1 Robinson approximations of dense graphs

The Robinson property for matrices was first introduced in the study of the classical seriation problem [81], whose objective is to order a set of items so that similar items are placed close to one another. The seriation question translates naturally into a question regarding symmetric matrices which can be turned into Robinson matrices. A symmetric matrix $A = [a_{ij}]$ is said to be a ***Robinson matrix*** if

$$i \leq j \leq k \implies a_{ik} \leq \min\{a_{ij}, a_{jk}\}. \tag{4.1}$$

A symmetric matrix A is called ***Robinsonian*** if it becomes a Robinson matrix after simultaneous application of a permutation π to its rows and columns. In that case, the permutation π is called a ***Robinson ordering*** of A. If the entries a_{ij} of the symmetric matrix A represent similarity of items i and j, then the Robinson ordering represents a linear arrangement of the items so that similar items are placed closer together. The problem of recognizing Robinsonian matrices and finding their Robinson orderings is well-studied, and can be solved in polynomial time (see e.g. [73]). However, what if the matrix in question is "almost" Robinsonian, and we are looking for an ordering which puts the matrix into an almost Robinson form? Answering this question turns out to be much more challenging; see for example [31, 17, 6] where different matrix norms have been used to model the distance between matrices.

A generic solution to the "almost Robinson approximation" problem involves two steps: (1) develop an algorithm to propose a suitable almost Robinson ordering, (2) estimate how close the reordered matrix is to being Robinson. This thesis looks entirely at step (2), focusing only on estimating how close in some norm a fixed matrix is from a Robinson matrix.

4.1.1 ℓ_1 approximations of matrices

We begin our discussion of Robinson approximations with the parameter Γ_1 that simply sums the normalized so-called "non-Robinson" parts of an $n \times n$ matrix. It was originally defined in [39] and we shall repeat its definition here, naturally extending its definition to include symmetric step functions w_n and recalling that the space of such functions is denoted $\mathcal{S}$. However, we first begin with the introduction of some notation to allow us to define Γ_1 in a more compact fashion.

Definition 4.1.1 (Cell average)**.** Let $w \in \mathcal{W}$ and let $A, B \subseteq [0,1]$ be measurable. The ***cell average*** of w over $A \times B$ is given by

$$\overline{w}(A \times B) = \frac{1}{|A \times B|} \iint_{A \times B} w \; dxdy. \tag{4.2}$$

With the notion of a cell average rigorously defined, we now present the definition of Γ_1 in the setting of matrices as well as the more general setting of step functions.

Definition 4.1.2 ([39])**.** Let A be a symmetric matrix of size n. Then

$$\Gamma_1(A) = \frac{1}{n^3} \sum_{1 \le i < k < j \le n} [A_{i,j} - A_{i,k}]_+ + [A_{i,j} - A_{k,j}]_+,$$

where $[x]_+ = \max(x, 0)$. Furthermore, let $[0,1]$ be partitioned into sets $I_1, \ldots, I_n$. Then for any step function $w_n \in \mathcal{S}$ with steps $I_i \times I_j$,

$$\Gamma_1(w_n) = \frac{1}{n^3} \sum_{1 \le i < k < j \le n} \Big[\overline{w}_n(I_i \times I_j) - \overline{w}_n(I_i \times I_k)\Big]_+ + \Big[\overline{w}_n(I_i \times I_j) - \overline{w}_n(I_k \times I_j)\Big]_+.$$

This parameter has many desirable properties, least of which is its ability to recognize Robinson matrices/step functions. Foremost among these properties is its computational requirements—it can obviously be computed in polynomial time, a trait none of the other parameters we introduce in this thesis share. Unfortunately, as we will later see, it has its own major failings which make it most suitable as a measurement tool for fixed matrices or step functions, rather than convergent sequences of such objects. We state the main approximation result concerning Γ_1 below.

Theorem 4.1.3 (Theorem 2.2, [39]). *Let A be an $n \times n$ symmetric matrix. Then there exists an $n \times n$ symmetric Robinson matrix R such that*

$$\|A - R\|_1 \leq 26\Gamma_1(A)^{\frac{1}{3}}.$$

Moreover, R can be computed in polynomial time. In addition, if A is binary, then there exists a binary matrix R satisfying the above conditions.

It is not necessarily the case that this bound gives the best possible approximation; however, there is also an easily attainable lower bound on this ℓ^1-difference that we state and prove below for completeness.

Lemma 4.1.4 (Lemma 2.3, [39]). *For every pair of symmetric $n \times n$ matrices A and R, if R is a Robinson matrix than*

$$\|A - R\|_1 \geq \frac{1}{4}\Gamma_1(A).$$

Thus, any Robinson approximation of A can do no better (in the ℓ^1-distance, at least) than $\frac{1}{4}\Gamma_1(A)$.

Proof. Since $[\cdot]_+$ is subadditive, it must be that

$$\Gamma_1(A) = \Gamma_1(A - R + R) \leq \Gamma_1(A - R) + \Gamma_1(R) = \Gamma_1(A - R),$$

as R is Robinson and so $\Gamma_1(R) = 0$. Moreover, for every symmetric $n \times n$ matrix B, it must be that

$$\begin{aligned}
\Gamma_1(B) &= \frac{1}{n^3} \sum_{1 \le i < k < j \le n} [B_{i,j} - B_{i,k}]_+ + [B_{i,j} - B_{k,j}]_+ \\
&\le \frac{1}{n^3} \sum_{1 \le i < k < j \le n} |B_{i,j}| + |B_{i,k}| + |B_{i,j}| + |B_{k,j}| \\
&\le \frac{4}{n^3} \sum_{k=1}^{n} \sum_{1 \le i,j \le n} |B_{i,j}| \le 4\|B\|_1.
\end{aligned}$$

We let $B = A - R$ to finish the proof. □

When applying Theorem 4.1.3 to binary matrices, one obtains an interesting corollary for graphs. If one edits the adjacency matrix A_G of a graph G so that the diagonal entries are all equal to 1, they will notice this augmented matrix is Robinson if and only if all of the 1-entries are consecutive along the rows and columns. This is known as the ***symmetric consecutive ones property*** and characterizes ***unit interval graphs*** or equivalently [83] ***proper interval graphs*** [21, 22, 65].

When Γ_1 is applied to an augmented adjacency matrix, it counts the number of triples (i, j, k) with $1 \le i < j < k \le n$ such that i is adjacent to k but j is adjacent to neither i nor k. However, the unnormalized ℓ^1-distance $\|A_G - A_H\|_1$ between the augmented adjacency matrices A_G and A_H of two labeled graphs G and H represents the ***edit distance*** between those graphs. This quantity, denoted by $ed(G, H)$, is the minimum number of edge deletions and additions that need to be performed on G to transform it in H. With this in mind, Theorem 4.1.3 directly implies the following corollary.

Corollary 4.1.5 (Corollary 2.4, [40])**.** *For every graph G on vertex set $V = \{1, 2, \ldots, n\}$, there exists a unit interval graph H on vertex set V such that*

$$\frac{ed(G, H)}{n^2} \le 26\Gamma_1(G)^{\frac{1}{3}}.$$

However, even with all of these positive attributes, Γ_1 has a glaring flaw. It cannot be used to analyze the behavior of convergent sequence of graphs, dense or sparse, as it is unfortunately not the case that Γ_1 is continuous on $\mathcal{S}$ with respect to the cut norm.

Proposition 4.1.6. Γ_1 *is not continuous with respect to the cut norm; that is, there exists a sequence of graphons* $w_n \in \mathcal{S}$ *and a graphon* $w \in \mathcal{S}$ *such that* $\|w_n - w\|_\square \to 0$ *but* $\Gamma_1(w_n) \not\to \Gamma_1(w)$.

Proof. Let G_n be the Erdős-Rényi random graph with parameter $\frac{1}{2}$ on n vertices; that is, each vertex is connected to each other vertex independently with probability $\frac{1}{2}$. We let w_n be the step function representation of the adjacency matrix A_{G_n}; it is a classic result of graph limit theory that $\|w_n - \frac{1}{2}\|_\square \to 0$ as $n \to \infty$, where the constant $\frac{1}{2}$ is viewed as a graphon. We shall show that $\Gamma_1(w_n) \not\to \Gamma_1(\frac{1}{2}) = 0$. To do this, we will make use of the uniform integrability of $\{\Gamma_1(w_n)\}$ (that this family of random variables is uniformly integrable is simple to check). Let $\mathcal{G} = \lim_{n\to\infty} \Gamma_1(w_n)$; if this quantity doesn't exist, then we are done. Supposing it does exist, it must be that $\mathsf{E}(\Gamma_1(w_n)) \to \mathsf{E}(\mathcal{G})$ by uniform integrability. If Γ_1 were continuous, then it would ensure that $\mathsf{E}(\mathcal{G}) = 0$. Thus we shall show that $\mathsf{E}(\mathcal{G}) \neq 0$ to complete our proof.

Consider each "column" of w_n, indexed by I_i, where the i-th column has $i-1$ elements above the diagonal. Then, we can show that

$$\begin{aligned}\mathsf{E}(\Gamma_1(w_n)) &= \frac{1}{n^3} \sum_{1\le i<k<j\le n} \mathsf{E}\Big(\Big[\overline{w}_n(I_i \times I_j) - \overline{w}_n(I_i \times I_k)\Big]_+ + \Big[\overline{w}_n(I_i \times I_j) - \overline{w}_n(I_k \times I_j)\Big]_+\Big) \\ &\ge \frac{1}{n^3} \sum_{1\le i<k<j\le n} \mathsf{E}\Big(\Big[\overline{w}_n(I_i \times I_j) - \overline{w}_n(I_i \times I_k)\Big]_+\Big), \qquad (4.3)\end{aligned}$$

where (4.3) is the average value of $\Gamma_1(w_n)$ calculated only over the columns of w_n. We note that the only situation in which a positive value is added to the sum (4.3) is when $\overline{w}_n(I_i \times I_j) = 1$ and $\overline{w}_n(I_i \times I_k) = 0$ for $1 \le i < k < j \le n$. Fixing i in such a case, we know that $\overline{w}_n(I_i \times I_j) = 1$ with probability $1/2$; furthermore, there are $j - i - 1$ cells of the form $I_i \times I_k$ between the cell $I_i \times I_j$ and the diagonal, each of which has

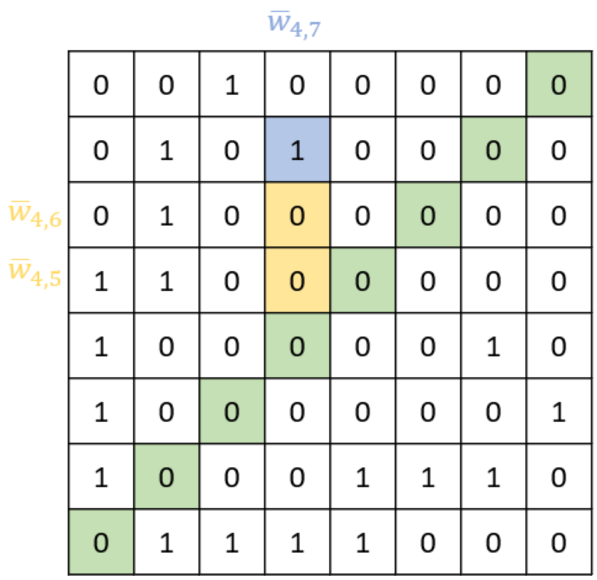

Figure 4.1: A graphon w that represents an 8x8 adjacency matrix where the diagonal is marked in green. The cell $I_4 \times I_7$, marked in blue, is being compared to the cells $I_4 \times I_6$ and $I_4 \times I_5$, both marked in yellow, in the same "column" of w.

value $\overline{w}_n(I_i \times I_k) = 0$ with probability $1/2$. For an illustration of how this comparison process works, please refer to Figure 4.1. Thus, it must be that

$$\frac{1}{n^3} \sum_{1 \leq i < k < j \leq n} \mathsf{E}\Big(\Big[\overline{w}_n(I_i \times I_j) - \overline{w}_n(I_i \times I_k)\Big]_+\Big) = \frac{1}{n^3} \sum_{1 \leq i < j \leq n} \frac{j-i-1}{4} = \frac{n^2 - 3n + 2}{24n^2}.$$

Therefore, it is clear that

$$\mathsf{E}(\mathcal{G}) = \lim_{n\to\infty} \mathsf{E}(\Gamma_1(w_n)) \geq \lim_{n\to\infty} \frac{n^2-3n+2}{24n^2} = \frac{1}{24},$$

and as this is clearly nonzero, Γ_1 cannot be continuous on $\mathcal{S}$. $\square$

To tackle the problem of a graph parameter that recognizes the Robinson property and is continuous with respect to the cut norm, we proceed to the next section.

4.1.2 Approximating in cut norm

In [19], the authors used a (labeled) graph parameter Γ^* to estimate how close a given graph's adjacency matrix is to a Robinson matrix in the case where proximity in the space of matrices was measured by cut norm. Indeed, they proved that the function Γ^* can be used to recognize graph sequences which are sampled from Robinson graphons. Recall that a graphon $w : [0,1]^2 \to \mathsf{R}$ is *Robinson* if

$$x \leq y \leq z \implies w(x,z) \leq \min\{w(x,y), w(y,z)\}. \tag{4.4}$$

We call a graphon ***Robinson almost everywhere***, or Robinson a.e. for short, if it is equal a.e. to a Robinson graphon. We recall the definition of Γ^*, noting that we have extended it to general matrices.

Definition 4.1.7 (The parameter Γ^*)**.** Let Q be a square matrix of size n, and I be a subset of $[n]$. We define

$$\Gamma^*(Q,I) := \frac{1}{n^3}\left(\sum_{j<k}\Big[\sum_{i\in I\cap\{1,\ldots,j\}}(Q_{i,k}-Q_{i,j})\Big]_+ + \sum_{j<k}\Big[\sum_{i\in I\cap\{k,\ldots,n\}}(Q_{i,j}-Q_{i,k})\Big]_+\right),$$

where $[x]_+ = \max\{x,0\}$. We define $\Gamma^*(Q)$ as $\Gamma^*(Q) := \sup_{I\subseteq[n]}\Gamma^*(Q,I)$.

Proposition 4.1.8 (Proposition 3.4, [19])**.** *A graph G is one dimensional geometric (equivalently, A_G is Robinson) if and only if $\Gamma^*(A_G) = 0$.*

Proof. Let G be a one dimensional geometric graph; then, there is an ordering of G such that A_G is a Robinson matrix. Thus, by definition, for $i<j<k$, $(A_G)_{i,k} \leq \min\{(A_G)_{i,j}, (A_G)_{j,k}\}$. Therefore, for all $I \subset [n]$,

$$\frac{1}{n^3}\sum_{j<k}\Big[\sum_{i\in I\cap\{1,\ldots,j\}}((A_G)_{i,k}-(A_G)_{i,j})\Big]_+ = 0$$
$$\frac{1}{n^3}\sum_{j<k}\Big[\sum_{i\in I\cap\{k,\ldots,n\}}((A_G)_{i,j}-(A_G)_{i,k})\Big]_+ = 0,$$

showing that $\Gamma^*(A_G) = 0$.

For the reverse direction, let $\Gamma^*(A_G) = 0$. Then for all subsets $I \subset [n]$ it must be that $\Gamma^*(A_G, I) = 0$. Thus, for $I = \{a\}$, it must be that

$$\frac{1}{n^3}\sum_{j<k}\Big[\sum_{i\in\{a\}\cap\{1,\dots,j\}}((A_G)_{i,k} - (A_G)_{i,j})\Big]_+ = 0$$
$$\frac{1}{n^3}\sum_{j<k}\Big[\sum_{i\in\{a\}\cap\{k,\dots,n\}}((A_G)_{i,j} - (A_G)_{i,k})\Big]_+ = 0,$$

which means that for all $a < j < k$, that $(A_G)_{a,k} \leq (A_G)_{a,j}$, and that for all $j < k < a$ that $(A_G)_{a,j} \leq (A_G)_{a,k}$. Therefore it must be the case that for $i < j < k$ that $(A_G)_{i,k} \leq \min\{(A_G)_{i,j}, (A_G)_{j,k}\}$, implying that A_G is Robinson. □

The function Γ was first defined in [19, Definition 4.1] for $w \in \mathcal{W}_0$, but it can be easily generalized to $\mathcal{W}^1$.

Definition 4.1.9 (The parameter Γ)**.** For $w \in \mathcal{W}^1$ and a measurable subset A of $[0,1]$, we define

$$\Gamma(w, A) := \iint_{y<z}\Big[\int_{x\in A\cap[0,y]}(w(x,z) - w(x,y))dx\Big]_+ dydz$$
$$+ \iint_{y<z}\Big[\int_{x\in A\cap[z,1]}(w(x,y) - w(x,z))dx\Big]_+ dydz,$$

where $[x]_+ = \max(x, 0)$. Moreover, $\Gamma(w)$ is defined as

$$\Gamma(w) := \sup_{A\in\mathcal{A}}\Gamma(w, A),$$

where the supremum is taken over all measurable subsets of $[0,1]$.

Like its discrete counterpart Γ^*, the parameter Γ also recognizes the Robinson property—even for L^p graphons. We display a short proof of this here alongside some relationships between Γ and the L^p norms.

Proposition 4.1.10. *Let $1 \leq p \leq \infty$, and suppose $w \in \mathcal{W}^p$. Then we have* *(i)* $|\Gamma(w)| \leq 4\|w\|_1 \leq 4\|w\|_p$.

(ii) w *is Robinson a.e. if and only if* $\Gamma(w) = 0$.

Proof. To prove (i), we recall the definition of $\Gamma(w)$ and apply the triangle inequality alongside the fact that for any real number x, we have $[x]_+ \leq |x|$, to show that

$$\begin{aligned}\Gamma(w, A) &\leq \iint_{y<z} \int_{x\in A\cap[0,y]} \Big(|w(x,z)| + |w(x,y)|\Big) dxdydz \\ &+ \iint_{y<z} \int_{x\in A\cap[z,1]} \Big(|w(x,z)| + |w(x,y)|\Big) dxdydz \\ &\leq 4\|w\|_1 \leq 4\|w\|_p,\end{aligned}$$

proving the claim.

The proof of (ii) is very similar to the proof of the analogous result for $\mathcal{W}_0$, but to be self-contained, we include a short proof here. We only prove that if $\Gamma(w) = 0$ then w must be a.e. Robinson; the converse direction is trivial. For $n \in \mathsf{N}$, consider the partition $\mathcal{P}_n = \{I_1, ..., I_n\}$, with $I_i = (\frac{i-1}{n}, \frac{i}{n}]$ for $1 < i \leq n$ and $I_1 = [0, \frac{1}{n}]$. Define $w_n := (w)_{\mathcal{P}_n}$, and let $\mathbb{1}_{i,j}$ denote the characteristic function of $I_i \times I_j$. As $w \in L^1([0,1]^2)$, the sequence of L^1 graphons $\{w_n\}_{n\in\mathsf{N}}$ satisfies $\|w_n - w\|_1 \to 0$ as $n \to \infty$, which therefore implies that there exists a subsequence $\{w_{n_k}\}_{k\in\mathsf{N}}$ converging a.e. to w. Thus it is enough to show that w_n is Robinson for all n, as a.e. pointwise limits of Robinson graphons are Robinson a.e. as well. Note that for $i < j < j+1 \leq n$,

$$\int_{y\in I_j} \int_{z\in I_{j+1}} \left[\int_{x\in I_i} w(x,z) - w(x,y)dx \right]_+ dydz \leq \Gamma(w, I_i) \leq \Gamma(w) = 0.$$

Thus for almost all $y \in I_j$ and $z \in I_{j+1}$,

$$\int_{x\in I_i} w(x,z)dx \leq \int_{x\in I_i} w(x,y)dx.$$

Integrating both sides over $y \in I_j$ and $z \in I_{j+1}$ implies that for $i < j < j+1 \leq n$, we have $w_n\mathbb{1}_{i,j+1} \leq w_n\mathbb{1}_{i,j}$. A similar argument shows that for $j < j+1 < i \leq n$, we have

$w_n \mathbb{1}_{i,j} \leq w_n \mathbb{1}_{i,j+1}$. This finishes the proof. □

So we have two parameters to measure the Robinson property: Γ^* and Γ. Let G be a labeled graph with n vertices and let w_G be its associated graphon. While it is natural to expect $\Gamma^*(G)$ and $\Gamma(w_G)$ to be similar, they are often not equal. However, the following result in [19] shows that they tend to the same value as n grows large.

Proposition 4.1.11 (Corollary 5.2, [19]). *Let G be a graph on n vertices and let w_G be the associated graphon of G with respect to a linear ordering $\prec$ of its vertices. Then*

$$\Gamma^*(G, \prec) = \Gamma(w_G) + \mathcal{O}(\frac{1}{n}).$$

We note also that Γ has the desirable property of cut norm continuity, which we state below in full. This makes it a suitable parameter for recognizing the Robinson property of dense graph sequences; with this in mind, we can use the similarity of Γ^* and Γ to develop a new parameter that respects the limit of dense graph sequences proper, rather than just their associated graphons.

Lemma 4.1.12 (Lemma 6.2, [19]). *Let $w_1, w_2 \in \mathcal{W}_0$. Then*

$$|\Gamma(w_1) - \Gamma(w_2)| \leq 2\|w_1 - w_2\| \;\; + 4\sqrt{\|w_1 - w_2\| \;\;}. \tag{4.5}$$

As Γ is continuous with respect to graphons in the cut norm, we hope that we can pass that continuity down to graph sequences. Namely, given a convergent sequence of dense graphs $\{G_n\}$ and a graphon $w \in \mathcal{W}_0$ where $\delta \;\; (G_n, w) \to 0$, we desire a result of the form

$$\Gamma^*(G_n) \to \Gamma(w). \tag{4.6}$$

Unfortunately, this does not hold in general. Γ is defined for a specific labeling and thus $\Gamma(w)$ is as well, ruining any hope of a limit result such as above being true. However, as the labeling is the only issue, we can simply define a new parameter that "divides out" the labeling and thus has a shot at attaining a result like (4.6). We define this new parameter below.

Definition 4.1.13. Let $w \in \mathcal{W}_0$. We define the new parameter $\widetilde{\Gamma}$ to be

$$\widetilde{\Gamma}(w) := \inf_{w' \approx w} \Gamma(w') = \inf\{\Gamma(w') : \delta\ (w, w') = 0\}.$$

It comes as no surprise that this new parameter is continuous with respect to δ ; indeed, this is the key ingredient that allows for a limit result such as (4.6) to hold with regards to $\widetilde{\Gamma}$.

Proposition 4.1.14 (Theorem 6.4, [19]). *Let $\{w_n\} \subset \mathcal{W}_0$ and let $w \in \mathcal{W}_0$ be such that $\delta\ (w_n, w) \to 0$. Then $\widetilde{\Gamma}(w_n) \to \widetilde{\Gamma}(w)$ as $n \to \infty$.*

This proposition immediately lends itself to the following corollary, showing our desired limit result.

Corollary 4.1.15 (Corollary 6.5, [19]). *Let $\{G_n\}$ be a sequence of simple graphs with $|V(G_n)| \to \infty$ and let $w \in \mathcal{W}_0$ such that $\delta\ (G_n, w) \to 0$. Then*

$$\Gamma^*(G_n) \to \widetilde{\Gamma}(w)$$

as $n \to \infty$.

It is worth pointing out that if a convergent graph sequence $\{G_n\}$ has limit w such that $\Gamma^*(G_n) \to 0$, the previous corollary states that $\widetilde{\Gamma}(w) = 0$. It must therefore be the case that there exist functions u with $\Gamma(u)$ arbitrarily small such that $\{G_n\}$ have similar homomorphism densities as the random graph $\mathsf{G}(n, u)$. It is thus tempting to conclude that the graphs $\{G_n\}$ are similar to random models sampled from a Robinson graphon; however, this does not follow from any of the previous results. Indeed, a δ -equivalence class may contain more than one Robinson graphon, so there is no concept of a "unique Robinson representation" of δ -equivalence class of graphons. It is also the case that there may be $u_1, u_2 \in \mathcal{W}_0$ such that $\delta_\square(w, u_1) = \delta_\square(w, u_2) = 0$ and $\widetilde{\Gamma}(w) = \widetilde{\Gamma}(u_1) = \widetilde{\Gamma}(u_2)$, but $\|u_1 - u_2\|_\square$ is large.

Facing these challenges, the authors in [19] ended their paper with this hopeful statement as a conjecture. They were proved right when this conjecture was resolved in the positive in [40], whose authors succeeded in a proof by developing a method to construct Robinson approximations to given graphons whose cut norm distance from the original was bounded by Γ. This construction, combined with a functional analytic argument, forms the backbone of their argument. We shall present these below, beginning with their construction.

Definition 4.1.16 (Robinson approximation)**.** Let $w \in \mathcal{W}_0$. R_w is called the ***Robinson approximation*** of w, and is defined as follows.

$$R_w(x, y) = R_w(y, x) = \sup\{\overline{w}(A \times B) : A \subseteq [0, x]; B \subseteq [y, 1]; |A| = |B| = \Gamma(w)^{2/7}\}, \tag{4.7}$$

where we define $\sup \emptyset = 0$.

This approximation is clearly Robinson, and we now present a result showing that it is indeed a good approximation in cut norm, omitting the proof due to its length.

Theorem 4.1.17 (Theorem 3.2, [40])**.** *Let $w \in \mathcal{W}_0$ and R_w be as defined in Definition 4.1.16. Then,*

$$\|R_w - w\|_\square \leq 14\Gamma(w)^{1/7}.$$

All that is left is to show that a convergent series of graphs $\{G_n\}$ such that $\Gamma(G_n) \to 0$ implies that G_n must converge to a Robinson graphon. To this end, [40] introduces the following definition, originally taken from matrix theory and then extended to graphons.

Definition 4.1.18 (Robinsonian graphons, [40])**.** A graphon $w \in \mathcal{W}_0$ is called ***Robinsonian*** if there exists a Robinson graphon $u \in \mathcal{W}_0$ such that $\delta_\square(w, u) = 0$.

Theorem 4.1.19 (Theorem 5.3, [40])**.** *Let $\{G_n\}$ be a growing sequence of graphs converging to a graphon $w \in \mathcal{W}_0$. Then w is Robinsonian if and only if $\widetilde{\Gamma}(G_n) \to 0$.*

We note this will imply our desired result, as if G_n is a graph sequence converging to w and $\Gamma(G_n) \to 0$, then it must be the case that $\widetilde{\Gamma}(G_n) \to 0$, implying that w is Robinsonian. Then there must exist some Robinson $u \in \mathcal{W}_0$ such that $\delta\ \ (w, u) = 0$ and then a simple application of the triangle inequality shows $\delta\ \ (G_n, u) \to 0$, implying that $\{G_n\}$ must converge to a Robinson graphon. It is thus clear that Γ is a suitable measurement of the Robinson property for graphons in the space $\mathcal{W}_0$. The question then becomes whether Γ can be extended to L^p graphons and enjoy similar properties. Proposition 4.1.10 shows that Γ identifies Robinson L^p graphons, in the sense that Γ attains 0 precisely when applied to Robinson L^p graphons. In order for Γ to be considered a suitable measurement for Robinson property of L^p graphons, it must satisfy continuity and stability properties similar to those in Lemma 4.1.12 and Theorem 4.1.17. The proof of Lemma 4.1.12 relies on continuity of triangular cut on $\mathcal{W}_0$. Namely, let χ be the characteristic function of the region in $[0,1]^2$ above the diagonal. Then the proof of Lemma 4.1.12 is based on the following two facts:

(a) $|\Gamma(w_1) - \Gamma(w_2)| \leq 2\|w_1 - w_2\|\ \ + 2\|(w_1 - w_2)\chi\|\ \$ for $w_1, w_2 \in \mathcal{W}^1$.

(b) $\|w\chi\|\ \ \leq 2\sqrt{\|w\|\ \ \|w\|_\infty}$ for $w \in \mathcal{W}^\infty$.

Thus, to determine the continuity of Γ on the space of L^p graphons, we need to study continuity of the triangular cut when applied to L^p graphons. From our results in Section 3.3, it follows that condition (b) cannot be generalized to $w \in \mathcal{W}^p$ with $1 < p < \infty$. In fact, we show that the map $M_\chi : w \mapsto w\chi$ is not $\|\cdot\|\ \ $-continuous on $\mathcal{W}^p$ for $1 < p < \infty$. So even though the definition of Γ can be extended to L^p graphons, the authors in [19] were not able to show that it forms a continuous function on these spaces.

4.2 Robinson approximations of sparse graphs

This section introduces a new function denoted by Λ, which can be used as a simpler and easier method to compute measurement for the Robinson property of L^p-graphons. We begin by recalling the following definitions.

Definition 4.2.1 (Set inequality). Let $A, B \subset \mathsf{R}$. We say that $A \leq B$ if for all $a \in A$ and $b \in B$, we have that $a \leq b$.

Having recalled how to order sets, we introduce our new parameter Λ below.

Definition 4.2.2 (Robinson parameter). Let $w \in \mathcal{W}^1$. Define

$$\Lambda(w) = \frac{1}{2} \sup_{\substack{A \leq B \leq C, \\ |A|=|B|=|C|}} \left[\iint_{A \times C} w \; dxdy - \iint_{B \times C} w \; dxdy \right] + \frac{1}{2} \sup_{\substack{X \leq Y \leq Z, \\ |X|=|Y|=|Z|}} \left[\iint_{X \times Z} w \; dxdy - \iint_{X \times Y} w \; dxdy \right] \tag{4.8}$$

where A, B, C and X, Y, Z are measurable subsets of $[0, 1]$.

Note that $\Lambda(w) \geq 0$ for all $w \in \mathcal{W}^1$, as we may take $A = B = C = X = Y = Z = \emptyset$. Moreover,

$$\Lambda(w) \leq \frac{1}{2} \left(\iint_{A \times C} |w| dxdy + \iint_{B \times C} |w| dxdy + \iint_{X \times Z} |w| dxdy + \iint_{X \times Y} |w| dxdy \right) \leq \|w\|_1,$$

implying that $\Lambda(w) \leq \|w\|_p$ for every $p \geq 1$. Informally, Λ can be thought of as extending the Robinson property from pointwise comparison to comparison of blocks. Λ can also be shown to be continuous with respect to the cut norm even for L^p graphons (see Proposition 4.2.3). This, alongside the fact that Λ requires only integration over blocks rather than more complicated operations, sets it aside from other similar graph parameters, which have not been shown to possess such continuity. We summarize basic properties of Λ below, but first introduce some definitions that will facilitate more understandable writing. The upper triangle of the unit square shall be denoted

$$\Delta = \left\{(x, y) \in [0,1]^2 : \ x \le y\right\}.$$

We borrow the following terminology from [40], recalling that the ***upper left (UL)*** and ***lower right (LR)*** regions of a given point $(a, b) \in \Delta$ are given by

$$\begin{aligned} \mathrm{UL}(a,b) &= [0,a] \times [b,1], \\ \mathrm{LR}(a,b) &= [a,b] \times [a,b] \cap \Delta. \end{aligned}$$

These regions prove useful to define as the behaviour of a Robinson graphon at some point (a, b) is intuitively determined by the value of the graphon in $\mathrm{UL}(a, b)$ and $\mathrm{LR}(a, b)$: the value at (a, b) must be bounded below by any value in $\mathrm{UL}(a, b)$ and bounded above by any value in $\mathrm{LR}(a, b)$. As is noted in [40], these regions can also be used to form an equivalent statement of the Robinson property, though we omit such a statement here.

Proposition 4.2.3. *Let $1 \le p \le \infty$. Suppose $w, u \in \mathcal{W}^p$. Then we have*

(i) *Λ is continuous with respect to cut-norm, i.e., $|\Lambda(w) - \Lambda(u)| \le 2\|w - u\|$.*

(ii) *Λ characterizes Robinson L^p-graphons. That is, w is Robinson a.e. if and only if $\Lambda(w) = 0$.*

(iii) *Λ is subadditive, i.e., $\Lambda(w + u) \le \Lambda(w) + \Lambda(u)$.*

Proof. To prove (i), let $w, u \in \mathcal{W}^1$. Then by definition of supremum, for all $\epsilon > 0$, there exist measurable sets $A \le B \le C$ with equal size and measurable sets $X \le Y \le Z$ with equal size such that

$$\frac{1}{2}\Big(\iint_{A\times C} w\ dxdy - \iint_{B\times C} w\ dxdy + \iint_{X\times Z} w\ dxdy - \iint_{X\times Y} w\ dxdy\Big) \ge \Lambda(w) - \epsilon \tag{4.9}$$

Thus, combining (4.9) with the definition of $\Lambda(u)$, we get that

$$\begin{aligned}
2(\Lambda(w) - \Lambda(u) - \epsilon) \;&\leq\; \iint_{A\times C} w\ dxdy - \iint_{B\times C} w\ dxdy + \iint_{X\times Z} w\ dxdy - \iint_{X\times Y} w\ dxdy \\
&-\; \iint_{A\times C} u\ dxdy + \iint_{B\times C} u\ dxdy - \iint_{X\times Z} u\ dxdy + \iint_{X\times Y} u\ dxdy \\
&=\; \iint_{A\times C} (w-u)dxdy + \iint_{B\times C} (u-w)dxdy \\
&+\; \iint_{X\times Z} (w-u)dxdy + \iint_{X\times Y} (u-w)dxdy \\
&\leq\; 2\|w-u\| \ + 2\|u-w\| \ = 4\|w-u\| \ ,
\end{aligned}$$

where the last inequality is due to the fact that $\iint_{A\times C}(w-u)dxdy \leq \|w-u\|$ by definition, with similar logic holding for the other three terms of the inequality. The proof for $\Lambda(u) - \Lambda(w)$ follows the same logic. Since $\epsilon > 0$ is arbitrary, taking $\epsilon \to 0$, we get $|\Lambda(w) - \Lambda(u)| \leq 2\|w-u\|$.

To prove (ii), observe that if w is Robinson a.e., then $\Lambda(w) \leq 0$. Taking A, B, C and X, Y, Z of measure 0, we get that $\Lambda(w) = 0$. To prove the reverse direction, suppose $\Lambda(w) = 0$. For $n \in \mathsf{N}$, let $w_n = (w)_{\mathcal{P}_n}$, where $\mathcal{P}_n$ is the partition $\{I_i = (\frac{i-1}{n}, \frac{i}{n}]\}_{i=1}^{n}$. Note that as $\Lambda(w) = 0$, for any choice of $A \leq B \leq C \subseteq [0,1]$ of equal size, we have

$$\iint_{A\times C} w(x,y)dxdy \leq \iint_{B\times C} w(x,y)dxdy$$

and for $X \leq Y \leq Z \subseteq [0,1]$ of equal size, we have

$$\iint_{X\times Z} w(x,y)dxdy \leq \iint_{X\times Y} w(x,y)dxdy.$$

Applying the above inequalities to the sets I_i, we observe that w_n is Robinson for each n. Since $w_n \to w$ in the L^1 norm, using the Borel-Cantelli Lemma, there exists a subsequence of $\{w_n\}$ that converges to w pointwise almost everywhere in $[0,1]^2$. Thus, we claim that w must be Robinson a.e. as well. To prove this, note that as $w_n \to w$ pointwise a.e., there exists some set $S \subset [0,1]^2$ such that $|S| = 0$ and

$w(x,y) = \lim_n w_n(x,y)$ for $(x,y) \in [0,1]^2 \setminus S$. As w is symmetric, we can assume that S is symmetric. We define the function

$$\widetilde{w}(y,x) = \widetilde{w}(x,y) = \begin{cases} \sup\limits_{(u,v)\in \mathrm{UL}(x,y)\setminus S} w(u,v) & (x,y) \in S \cap \Delta \\ w(x,y) & (x,y) \notin \Delta \setminus S \end{cases}.$$

It is clear that $w = \widetilde{w}$ a.e.; we also claim that $\widetilde{w}$ is Robinson. Consider a point $(x,y) \in \Delta$. We wish to compare $\widetilde{w}(x,y)$ to the value of $\widetilde{w}$ over the set $\mathrm{UL}(x,y)$. This breaks down into two cases.

Case 1: $(x,y) \in S$. In this case, $\widetilde{w}(x,y)$ is greater than or equal to any value of $\widetilde{w}$ in $\mathrm{UL}(x,y)$ by definition of supremum.

Case 2: $(x,y) \notin S$. Let us first consider a point $(x_1,y_1) \in \mathrm{UL}(x,y) \setminus S$. This implies that $\widetilde{w}(x_1,y_1) = w(x_1,y_1) = \lim_n w_n(x_1,y_1)$. Therefore, $\widetilde{w}(x,y) \geq \widetilde{w}(x_1,y_1)$, as $w_n(x,y) \geq w_n(x_1,y_1)$ for all n. Next, we consider a point $(x_2,y_2) \in \mathrm{UL}(x,y) \cap S$. Here, we note that for any $\epsilon > 0$, there exists a point $(u,v) \in \mathrm{UL}(x_2,y_2) \setminus S$ such that $\widetilde{w}(u,v) \geq \widetilde{w}(x_2,y_2) - \epsilon$. However, as (u,v) is not in S, $w_n(u,v) \leq w_n(x,y)$ for all n, implying that $\widetilde{w}(x,y) = w(x,y) \geq w(u,v) = \widetilde{w}(u,v) \geq \widetilde{w}(x_2,y_2) - \epsilon$. Letting $\epsilon \to 0$, it is clear that $\widetilde{w}(x,y) \geq \widetilde{w}(x_2,y_2)$, implying that $\widetilde{w}(x,y)$ is greater than or equal to any value of $\widetilde{w}$ in $\mathrm{UL}(x,y)$.

Thus, $\widetilde{w}$ must be Robinson, and as $w = \widetilde{w}$ a.e., it must be the case that w is Robinson a.e., proving the claim.

To prove item (iii), note that

$$\begin{aligned}
\Lambda(w+u) &= \frac{1}{2} \sup_{\substack{A\leq B\leq C,\\ |A|=|B|=|C|}} \left[\iint_{A\times C} (w+u)dxdy - \iint_{B\times C} (w+u)dxdy \right] \\
&+ \frac{1}{2} \sup_{\substack{X\leq Y\leq Z,\\ |X|=|Y|=|Z|}} \left[\iint_{X\times Z} (w+u)dxdy - \iint_{X\times Y} (w+u)dxdy \right] \\
&\leq \frac{1}{2} \sup_{\substack{A\leq B\leq C,\\ |A|=|B|=|C|}} \left[\iint_{A\times C} w\ dxdy - \iint_{B\times C} w\ dxdy \right]
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{2}\sup_{\substack{X\le Y\le Z,\\ |X|=|Y|=|Z|}}\left[\iint_{X\times Z} w\ dxdy - \iint_{X\times Y} w\ dxdy\right]\\
&+\frac{1}{2}\sup_{\substack{A\le B\le C,\\ |A|=|B|=|C|}}\left[\iint_{A\times C} u\ dxdy - \iint_{B\times C} u\ dxdy\right]\\
&+\frac{1}{2}\sup_{\substack{X\le Y\le Z,\\ |X|=|Y|=|Z|}}\left[\iint_{X\times Z} u\ dxdy - \iint_{X\times Y} u\ dxdy\right]\\
&=\Lambda(w)+\Lambda(u),
\end{aligned}
$$

finishing the proof. □

Remark 4.2.4. We draw attention to a particular property of Λ: Let $S_u \le S_l \le T_l \le T_u$ be subsets of $[0,1]$ such that $|S_u| = |S_l| = |T_l| = |T_u| = \alpha$. Then

$$
\Lambda(w) \ge \frac{\alpha^2}{2}(\overline{w}(S_u\times T_u) - \overline{w}(S_l\times T_l)).
$$

To prove this, let $A = S_u$, $B = S_l$, $C = T_u$, $X = S_l$, $Y = T_l$, and $Z = T_u$. Then by definition,

$$
\begin{aligned}
\Lambda(w) &\ge \frac{1}{2}\left(\iint_{S_u\times T_u} w\ dxdy - \iint_{S_l\times T_u} w\ dxdy + \iint_{S_l\times T_u} w\ dxdy - \iint_{S_l\times T_l} w\ dxdy\right)\\
&= \frac{1}{2}\left(\iint_{S_u\times T_u} w\ dxdy - \iint_{S_l\times T_l} w\ dxdy\right)\\
&= \frac{\alpha^2}{2}(\overline{w}(S_u\times T_u) - \overline{w}(S_l\times T_l)),
\end{aligned}
$$

proving the claim. This, alongside the fact that Λ is continuous with respect to the cut norm and recognizes Robinson graphons, classifies it as a "Γ-type function" as defined in [40]. These were shown to have stable Robinson approximations for bounded graphons and to recognize when limits of convergent graph sequences are Robinson. While stability results for Λ already exist for bounded graphons, we prove here much better estimates than those given in [40] due to the improved definition of Λ.

4.2.1 Approximating L^p graphons

Importantly, we also show that Λ is stable when regarded as a parameter on L^p graphons for $p > 13/3$. We begin by defining the Robinson approximation of an L^p graphon.

Definition 4.2.5 (Robinson approximation for graphons)**.** Given a graphon $w \in \mathcal{W}^p$ where $p \geq 1$ and a fixed parameter $0 < \alpha < 1$, the α-*Robinson approximation* R_w^α of w is defined as follows: For all $(x, y) \in \Delta$,

$$R_w^\alpha(x,y) = R_w^\alpha(y,x) = \sup\left\{\overline{w}(S \times T) \,:\, S \times T \subseteq \mathrm{UL}(x,y),\, |S| = |T| = \alpha\right\}, \quad (4.10)$$

where S, T are measurable and taking the convention that $\sup \emptyset = 0$. Moreover, we say that if $\alpha = 0$, then $R_w^\alpha = w$.

Clearly, if w is non-negative, then R_w^α is a Robinson graphon. Also, it is simple to verify that R_w^α attains 0 on every point with distance at most α to the boundary of the square. We summarize some other properties of R_w^α below.

Lemma 4.2.6. *Given graphons $w, u \in \mathcal{W}^p$ where $p \geq 1$ and a xed parameter $0 < \alpha < 1$, we have:*

(i) If $u, w \in \mathcal{W}^p$ such that $u \leq w$ pointwise, then $0 \leq R_w^\alpha - R_u^\alpha \leq R_{w-u}^\alpha$.

(ii) If $w \in \mathcal{W}^p$, then $\|R_w^\alpha\|_\infty \leq \alpha^{-\frac{2}{p}} \|w\|_p$.

(iii) If $\{w_n\} \subseteq \mathcal{W}^p$ such that $\|w_n - w\|_1 \to 0$, then $\|R_w^\alpha - R_{w_n}^\alpha\| \to 0$ as $n \to \infty$.

We note that for item (iii), only convergence in the L^1 norm | the weakest norm | is required. Therefore, convergence in any L^p norm is enough to guarantee the convergence of the α-Robinson approximations in the cut norm.

Proof. To prove (i), let $(x, y) \in \Delta$. By definition of the α-Robinson approximation, for every $\epsilon > 0$, there exist sets $A, B \subseteq \mathrm{UL}(x, y)$ such that $R_w^\alpha(x, y) \geq \overline{w}(A \times B) \geq R_w^\alpha(x, y) - \epsilon$ and $C, D \subseteq \mathrm{UL}(x, y)$ such that $R_u^\alpha(x, y) \geq \overline{u}(C \times D) \geq R_u^\alpha(x, y) - \epsilon$. We also note that $|A| = |B| = |C| = |D| = \alpha$. Consider that

$$0 \leq \frac{1}{\alpha^2} \iint_{C \times D} (w - u)\,dxdy$$

$$
\begin{aligned}
&= \frac{1}{\alpha^2}\iint_{C\times D} w \ dxdy - \frac{1}{\alpha^2}\iint_{C\times D} u \ dxdy \\
&\leq R_w^\alpha(x,y) - R_u^\alpha(x,y) + \epsilon,
\end{aligned}
$$

showing the left hand side of the original claim when $\epsilon \to 0$. For the right hand side, we note that

$$
\begin{aligned}
R_w^\alpha(x,y) - \epsilon - R_u^\alpha(x,y) &\leq \frac{1}{\alpha^2}\iint_{A\times B} w \ dxdy - \frac{1}{\alpha^2}\iint_{C\times D} u \ dxdy \\
&\leq \frac{1}{\alpha^2}\iint_{A\times B} (w-u) dxdy \\
&\leq \sup_{\substack{A,B\subseteq \mathrm{UL}(x,y)\\ |A|=|B|=\alpha}} \frac{1}{\alpha^2}\iint_{A\times B} (w-u) dxdy = R_{w-u}^\alpha(x,y),
\end{aligned}
$$

showing that (i) holds true, again by letting $\epsilon \to 0$.

To prove (ii), we note that for every $\epsilon > 0$, there exist sets $A, B \subseteq \mathrm{UL}(x,y)$ where $|A| = |B| = \alpha$ such that $R_w^\alpha(x,y) \geq \overline{w}(A\times B) \geq R_w^\alpha(x,y) - \epsilon$. Then, by Hölder's inequality,

$$
R_w^\alpha(x,y) - \epsilon \leq \frac{1}{\alpha^2}\iint_{A\times B} w \ dxdy \leq \frac{1}{\alpha^2}\|w\|_p \|\mathbb{1}_{A\times B}\|_q = \alpha^{-\frac{2}{p}}\|w\|_p,
$$

where $\|\mathbb{1}_{A\times B}\|_q = \alpha^{2/q} = \alpha^{2-2/p}$. Thus, letting $\epsilon \to 0$, we arrive at the claim of (ii).

To prove (iii), we shall define an auxiliary function $l_n(x,y) = \min(w_n(x,y), w(x,y))$, noting that it converges to w in L^1 norm. To prove this, we define the set $L_n = \{(x,y) \in [0,1]^2 : w_n(x,y) \leq w(x,y)$. We then consider that

$$
\begin{aligned}
\|l_n - w\|_1 &= \iint_{[0,1]^2} |l_n(x,y) - w(x,y)| dxdy \\
&= \iint_{[0,1]^2} |\min(w_n(x,y), w(x,y)) - w(x,y)| dxdy \\
&= \iint_{L_n^C} |w(x,y) - w(x,y)| dxdy + \iint_{L_n} |w_n(x,y) - w(x,y)| dxdy \\
&= \iint_{L_n} |w_n(x,y) - w(x,y)| dxdy \leq \iint_{[0,1]^2} |w_n(x,y) - w(x,y)| dxdy
\end{aligned}
$$

$$= \|w_n - w\|_1 \to 0.$$

With this knowledge in mind, we can use parts (i) and (ii) (specifically using the case $p = 1$) to get that

$$\begin{aligned}
\|R^\alpha_{w_n} - R^\alpha_w\| &\leq \|R^\alpha_{w_n} - R^\alpha_{l_n}\| \quad + \|R^\alpha_w - R^\alpha_{l_n}\| \\
&\leq \|R^\alpha_{w_n - l_n}\| \quad + \|R^\alpha_{w - l_n}\| \\
&\leq \|R^\alpha_{w_n - l_n}\|_\infty + \|R^\alpha_{w - l_n}\|_\infty \\
&\leq \alpha^{-2}\Big(\|w_n - l_n\|_1 + \|w - l_n\|_1\Big) \\
&\leq \alpha^{-2}\Big(\|w_n - w\|_1 + 2\|w - l_n\|_1\Big) \to 0,
\end{aligned}$$

completing the proof. □

Furthermore, though R^α_w is certainly not always continuous on $[0, 1]^2$—the outer boundary of thickness α of the unit square is set to 0 by definition—it is continuous on $[\alpha, 1 - \alpha]^2$.

Proposition 4.2.7. *Let $w \in \mathcal{W}^p$, $1 \leq p \leq \infty$, and $\alpha \in (0, 1)$. Then R^α_w is continuous on the set $[\alpha, 1 - \alpha]^2$.*

Proof. Let $\{(x_n, y_n)\} \subset [\alpha, 1-\alpha]^2$ be a sequence such that $(x_n, y_n) \to (x, y) \in [\alpha, 1-\alpha]^2$, where the standard Euclidean distance is used to consider convergence. Then, by definition of UL, it must be the case that

$$\mathrm{UL}(\min(x_n, x), \max(y_n, y)) \subseteq \mathrm{UL}(x, y) \subseteq \mathrm{UL}(\max(x_n, x), \min(y_n, y)). \tag{4.11}$$

By the definition of R^α_w, (4.11) implies that

$$R^\alpha_w(\min(x_n, x), \max(y_n, y)) \leq R^\alpha_w(x_n, y_n) \leq R^\alpha_w(\max(x_n, x), \min(y_n, y)). \tag{4.12}$$

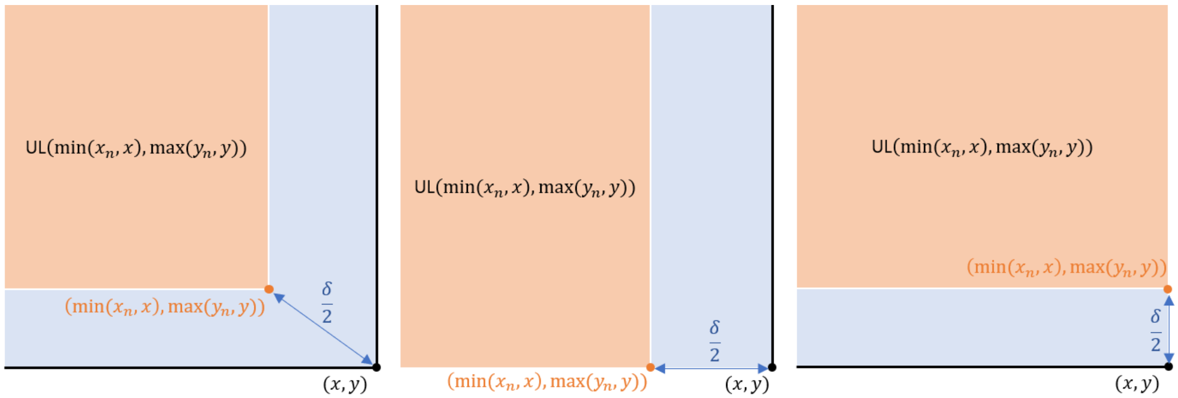

Figure 4.2: The blue area represents the symmetric difference between $\mathrm{UL}(x, y)$ and $\mathrm{UL}(\min(x_n, x), \max(y_n, y))$.

We shall show that both the upper and lower bound in (4.12) converge to $R_w^\alpha(x, y)$; by the Squeeze Theorem, this will imply our original claim. To do so, we let $\epsilon > 0$ and recall from elementary analysis that as w is a measurable function, for every $\epsilon > 0$, there exists a $\delta > 0$ such that if $S \subset [\alpha, 1-\alpha]^2$ and $|S| < \delta$, it must be that $\iint_S |w| dxdy < \epsilon$. Let $N \in \mathsf{N}$ be such that for all $n \geq N$, we have that $d((x_n, y_n), (x, y)) < \frac{\delta}{2}$. Defining the set $S := \mathrm{UL}(\min(x_n, x), \max(y_n, y)) \triangle \mathrm{UL}(x, y)$, it must then be that

$$R_w^\alpha(x, y) \leq R_w^\alpha(\min(x_n, x), \max(y_n, y)) + \iint_S |w| dxdy \leq R_w^\alpha(\min(x_n, x), \max(y_n, y)) + \epsilon,$$

as the measure of the symmetric difference of $\mathrm{UL}(\min(x_n, x), \max(y_n, y))$ and $\mathrm{UL}(x, y)$ is less than or equal to δ. For an explanation of this we refer to Figure 4.2, noting that the blue regions in each of the three possible cases cannot exceed $\frac{\delta}{2}(x+1-y)$ in measure. We recall that by definition it must be that $R_w^\alpha(\min(x_n, x), \max(y_n, y)) \leq R_w^\alpha(x, y)$, so

$$|R_w^\alpha(\min(x_n, x), \max(y_n, y)) - R_w^\alpha(x, y)| < \epsilon.$$

As this holds for a general $\epsilon > 0$, we let $\epsilon \to 0$ to get that $\lim_n R_w^\alpha(\min(x_n, x), \max(y_n, y)) = R_w^\alpha(x, y)$. A near identical argument can be performed to show that the similar result $\lim_n R_w^\alpha(\max(x_n, x), \min(y_n, y)) = R_w^\alpha(x, y)$ holds, the only difference being that the measure of the set $|\mathrm{UL}(\max(x_n, x), \min(y_n, y)) \triangle \mathrm{UL}(x, y)|$ is shown to be arbitrarily small. This is because that set is less than in measure than $\frac{\delta}{2}(x_n + 1 - y_n) \le \delta$ and so δ can once again be chosen small enough to carry on with the proof. Thus, by the Squeeze Theorem, it must be that $\lim_n R_w^\alpha(x_n, y_n) = R_w^\alpha(x, y)$, showing that R_w^α is continuous on $[\alpha, 1-\alpha]^2$. □

4.2.2 Stability of L^p approximations

In this section we present our main results on Λ. Because Λ functions as a measure of the Robinson property, it is desirable for a graphon with a small Λ value to be in some way ***close*** to a Robinson graphon. We call such a property ***Robinson stability*** and define it rigorously below.

Definition 4.2.8 (Robinson stability)**.** A graphon parameter Ψ is ***Robinson stable*** if there exists a decreasing function $f : [0, \infty) \to [0, \infty)$ with $\lim_{x \to 0^+} f(x) = 0$ satisfying the following: for any $w \in \mathcal{W}^p$, there exists a Robinson $R_w \in \mathcal{W}^p$ such that

$$\|w - R_w\| \le f(\Psi(w)).$$

We shall prove that Λ is Robinson stable; however, some more definitions are required before we begin. Our proof is very similar to the proof of Theorem 3.2 in [40] and so we shall use the same notation. Namely, given a graphon w such that $w \ge 0$ and $\lceil \|w\|_\infty \rceil = M$, we split Δ into smaller regions of three types—black, white, and grey regions—defined by the size of $\overline{w}$, given in Definition 4.1.1. Let $\alpha \in (0, 1)$ be as chosen in Definition 4.2.5 and let m be a fixed integer. For $k \in \{1, \ldots, mM - 1\}$, define the k-th black region $\mathcal{B}_k$, the k-th white region $\mathcal{W}_k$ and the k-th grey region $\mathcal{G}_k$ as follows.

$$\mathcal{B}_k = \Big\{(x, y) \in \Delta : \ x = y \ \text{ or } \ \exists\, S \times T \subseteq \mathrm{UL}(x, y) \text{ with } |S| = |T| = \alpha \text{ and } \overline{w}(S \times T) > \frac{k}{m}\Big\},$$
$$\mathcal{W}_k = \Big\{(x, y) \in \Delta \setminus \mathcal{B}_k : \ \exists\, S \times T \subseteq \mathrm{LR}(x, y) \text{ with } |S| = |T| = \alpha \text{ and } \overline{w}(S \times T) \le \frac{k}{m}\Big\},$$
$$\mathcal{G}_k = \Delta \setminus (\mathcal{B}_k \cup \mathcal{W}_k).$$

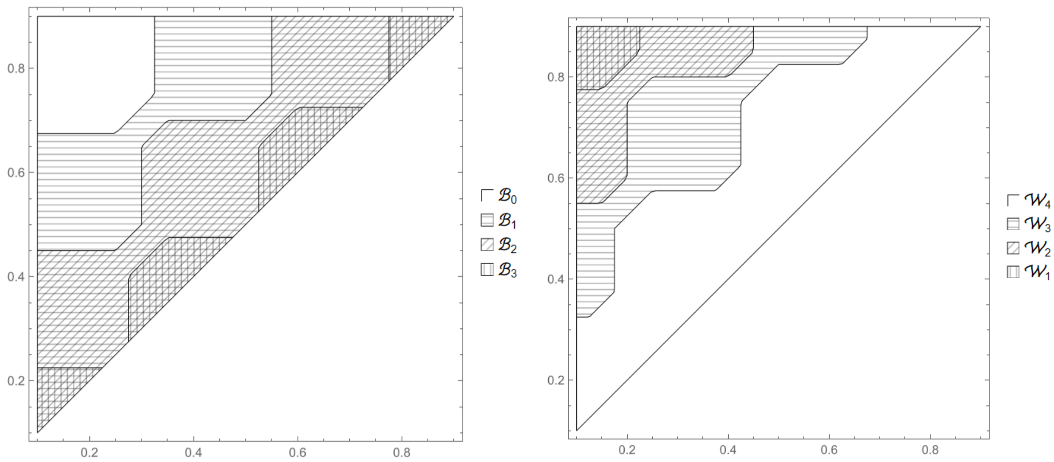

Figure 4.3: An example of black and white regions for $m = 4$ and $\alpha = 0.1$.

We set $\mathcal{B}_0 = \mathcal{W}_{mM} = \Delta$ and $\mathcal{W}_0 = \mathcal{B}_{mM} = \emptyset$ and let $\mathcal{G} = \bigcup_{k=1}^{mM-1} \mathcal{G}_k$. In addition, we also define the regions $\mathcal{R}_k := \mathcal{B}_k \cap \mathcal{W}_{k+1}$, though we do not give them a pithy name. Black regions provide lower bounds on the value of $\overline{w}$ while white regions provide upper bounds on $\overline{w}$. Grey regions provide no information about what values $\overline{w}$ could take on while the regions $\mathcal{R}_k$ are used to form tight "bands" on the behaviour of $\overline{w}$. Next, we observe that the $\mathcal{R}_k$ regions partition $\Delta \setminus \mathcal{G}$.

Lemma 4.2.9. *Let $w \in \mathcal{W}^\infty$ be such that $w \geq 0$ and $\lceil \|w\|_\infty \rceil = M$, and let $\mathcal{B}_k$, $\mathcal{W}_k$, and $\mathcal{G}_k$ be as de ned above. Then,*

$$\Delta \setminus \left(\bigcup_{k=1}^{mM-1} \mathcal{G}_k \right) = \bigcup_{k=0}^{mM-1} \mathcal{R}_k.$$

Proof. We first observe that

$$\Delta \setminus \left(\bigcup_{k=1}^{mM-1} \mathcal{G}_k \right) = \Delta \setminus \left(\bigcup_{k=1}^{mM-1} \Delta \setminus (\mathcal{B}_k \cup \mathcal{W}_k) \right) = \bigcap_{k=1}^{mM-1} (\mathcal{B}_k \cup \mathcal{W}_k).$$

Next, we consider the expansion of $(\mathcal{B}_1 \cup \mathcal{W}_1) \cap (\mathcal{B}_2 \cup \mathcal{W}_2) \cap \ldots \cap (\mathcal{B}_{mM-1} \cup \mathcal{W}_{mM-1})$ into expressions $X_1 \cap \ldots \cap X_{mM-1}$ with $X_i \in \{\mathcal{B}_i, \mathcal{W}_i\}$. We further note that $X_1 \cap \ldots \cap$

$X_{mM-1} = \emptyset$ whenever $X_i = \mathcal{W}_i$ and $X_j = \mathcal{B}_j$ for some $i < j$; thus, every nonempty term $X_1 \cap \ldots \cap X_{mM-1}$ from the above expansion must be of one of the following forms:

(i) $X_1 \cap \ldots \cap X_{mM-1} = \mathcal{B}_j \cap \mathcal{W}_{j+1} = \mathcal{R}_j$ with $1 \leq j < mM - 1$ if there is at least one black and one white region amongst the X_i.

(ii) $X_1 \cap \ldots \cap X_{mM-1} = \mathcal{W}_1 \cap \ldots \cap \mathcal{W}_{mM-1} = \mathcal{W}_1$ if all X_i are white.

(iii) $X_1 \cap \ldots \cap X_{mM-1} = \mathcal{B}_1 \cap \ldots \cap \mathcal{B}_{mM-1} = \mathcal{B}_{mM-1}$ if all X_i are black.

This completes the proof, as $\mathcal{W}_1 = \mathcal{W}_1 \cap \mathcal{B}_0 = \mathcal{R}_0$ and $\mathcal{B}_{mM-1} = \mathcal{B}_{mM-1} \cap \mathcal{W}_{mM} = \mathcal{R}_{mM-1}$. □

Remark* 4.2.10.** By definition, for every $k \in \mathbb{Z}^+$, the region $\mathcal{G}_k$ is bounded between ***upper and ***lower boundary*** functions $f_k, g_k : [0,1] \to [0,1]$, where f_k is the upper boundary of $\mathcal{B}_k$ and g_k is the lower boundary of $\mathcal{W}_k$. These are defined in [40] as follows:

$$f_k(x) = \sup\{z \in [x,1] : (x,z) \in \mathcal{B}_k\} \tag{4.13}$$

$$g_k(x) = \inf\{z \in [x,1] : (x,z) \in \mathcal{W}_k\} \tag{4.14}$$

where we set $\inf \emptyset = 1$ if it appears in the definition of g_k. Additionally, we define $f_0(x) = 1$ and $g_{mM}(x) = x$ for all $x \in [0,1]$ to represent the corresponding boundaries for $\mathcal{B}_0 = \mathcal{W}_{mM} = \Delta$. Finally, since f_k and g_{k+1} are the upper and lower boundaries of $\mathcal{B}_k$ and $\mathcal{W}_{k+1}$ respectively, if the region $\mathcal{R}_k$ is nonempty, then it is bounded from below by g_{k+1} and from above by f_k. We refer to Figure 4.3 for a visual representation of the regions and boundary curves previously mentioned.

Definition 4.2.11 (Boundary curves and crossing boundaries)**.** From the definition of the black and white regions, it is easy to see that the functions f_k and g_k are both increasing functions. So, they can only admit jump discontinuities. We naturally extend the graph of these functions to ***boundary curves*** by adding vertical line segments connecting any such discontinuities, denoting the resulting curves once again by f_k and g_k respectively.

Let $S, T \subseteq [0,1]$ be measurable. We say that $S \times T$ *crosses* a boundary curve f_k or g_k if the top-left corner of the cell is strictly above the boundary curve and its bottom-right corner is strictly below the curve. This definition is used as the set $S \times T$ is not necessarily a connected subset of $\mathbb{R}^2$ and so a boundary curve can go through the cell without having to intersect with it. See Figure 4.4 for an example of this phenomenon.

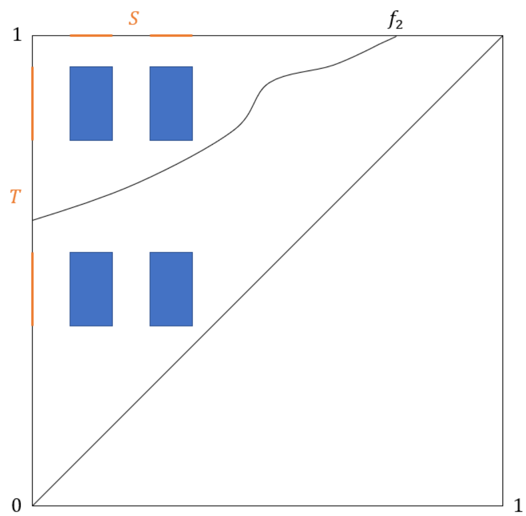

Figure 4.4: An example of a cell $S \times T$ that crosses a boundary curve f_2 without intersecting it.

Grey regions present a problem for predicting the value of R_w^α and therefore understanding how they behave is paramount. From the following lemma—whose proof is an adaptation of the proof of [40, Lemma 4.9]—it is clear that they cannot be too large in size or measure.

Lemma 4.2.12. *Let $k \in \mathbb{Z}^{\geq 0}$, $w \in \mathcal{W}^p$ with $w \geq 0$, and $\alpha \in (0,1)$. Then, $\overline{\mathcal{G}_k}$ does not contain any $\beta \times \beta$ square, where $\beta > \alpha$. Here, $\overline{\mathcal{G}_k}$ denotes the closure of $\mathcal{G}_k$ in the Euclidean topology of $\mathbb{R}^2$.*

The proof of our main theorem is based on the idea that the total area of all the grey regions is small. This allows us to concentrate on the behaviour of w and

R_w^α inside the regions $\mathcal{R}_k$, where their values are strictly controlled. We show that their local average difference in these regions can be controlled by $\Lambda(w)$, leading to the conclusion that $\|R_w^\alpha - w\|_\square$ must be small.

We introduce two more lemmas here, each necessary for controlling the size of integrals over certain sets in the proof of our main result. We will present the proofs of Lemma 4.2.12, Lemma 4.2.13 and Lemma 4.2.14 in Section 4.2.4.

Lemma 4.2.13. *Let $u \in L^\infty([0,1])$ (not necessarily non-negative), and $P \subseteq [0,1]$ be a measurable subset such that $\int_P u \neq 0$. Let $0 < \beta < |P|$ be fixed. Then P can be partitioned into $N := \lceil |P|/\beta \rceil$ subsets $P_1, \ldots, P_N$ so that the following conditions are satisfied:*

(i) $P_1 \leq \ldots \leq P_{N-1}$.

(ii) $|P_i| = \beta$ *for* $1 \leq i \leq N-1$ *and* $|P_N| \leq \beta$.

(iii) $\left|\int_{P_N} u\right| \leq \frac{1}{N}\left|\int_P u\right|$.

Lemma 4.2.14. *Let $f \in L^1([0,1]^2)$, and let $S, S' \subseteq [0,1]$ be measurable subsets such that $|S| = |S'|$. Suppose for a constant $C > 0$ we have*

$$\iint_{S\times S'} f \; dxdy \geq C.$$

Then, for every $\alpha \in (0,1)$, there exist measurable sets $T \subset S$ and $T' \subset S'$ such that $|T| = |T'| = \alpha|S|$ and

$$\frac{1}{|T\times T'|}\iint_{T\times T'} f \; dxdy \geq \frac{C}{|S\times S'|}.$$

Prefacing our main result, we begin first with a necessary proposition—as our main proof heavily features "cutting" graphons off at certain values, we require a way to control their behaviour once cut. The following proposition provides us with such a way.

Proposition 4.2.15. *Let $p > 2$ and $w \in \mathcal{W}^\infty$ with $w \geq 0$. Suppose $\|w\|_p = 1$. If R_w^α is the Robinson approximation of w with parameter $\alpha = \|w\|_\infty^{-\frac{p}{3p-2}} \Lambda(w)^{\frac{2p}{5p-2}}$, then*

$$\|R_w^\alpha - w\|_\square \leq 6\left(1 + (8\|w - R_w^\alpha\|_p + 2)\|w\|_\infty^{\frac{2p}{3p-2}}\right)\Lambda(w)^{\frac{p-2}{5p-2}}. \tag{4.15}$$

The story of this proof is as follows: We begin by locating a set $S \times T$ such that the value of $w - R_w^\alpha$ over that set is at least half its cut norm. Then we show that there is at least one $\alpha \times \alpha$ cell inside $(S \times T) \cap \Delta$, which is entirely contained inside some $\mathcal{R}_k$. This allows us to make tight estimates on the size of both w and R_w^α inside this subset due to the definition of $\mathcal{R}_k$; using these estimates alongside the inequality in Remark 4.2.4, we arrive at the statement of our proposition.

Proof. Let $M := \lceil \|w\|_\infty \rceil$. By [66, Lemma 8.10], there exist measurable $S, T \subseteq [0,1]$ so that

$$\left| \iint_{S\times T} (R_w^\alpha - w) dx dy \right| = \|R_w^\alpha - w\|_\square .$$

Replacing $S \times T$ with $T \times S$ if necessary, we can assume without loss of generality that

$$\left| \iint_{(S\times T)\cap\Delta} (R_w^\alpha - w) dx dy \right| \geq \frac{1}{2}\|R_w^\alpha - w\|_\square . \tag{4.16}$$

Fix $\beta > \alpha$. Next, we split S into $N_1 := \lceil |S|/\beta \rceil$ subsets $S_1, S_2, \ldots, S_{N_1}$ so that the following conditions are satisfied:

(i) $S_1 \leq \ldots \leq S_{N_1-1}$.

(ii) $|S_i| = \beta$ for $1 \leq i \leq N_1 - 1$ and $|S_{N_1}| \leq \beta$.

(iii) $\left| \iint_{S_{N_1}\times T} (R_w^\alpha - w) dx dy \right| \leq \frac{\|R_w^\alpha - w\|_\square}{N_1}$.

To prove that such a partition of S exists, we apply Lemma 4.2.13 to the function $u(\cdot) = \int_T (R_w^\alpha - w)(\cdot, y) dy$ and $P = S$. Likewise, we split T into $N_2 = \lceil |T|/\beta \rceil$ subsets satisfying $T_1 \leq T_2 \leq \cdots \leq T_{N_2-1}$ with $|T_1| = |T_2| = \cdots = |T_{N_2-1}| = \beta$ and $|T_{N_2}| \leq \beta$, so that $\left| \iint_{S\times T_{N_2}} (w - R_w^\alpha) dx dy \right| \leq \frac{\|R_w^\alpha - w\|_\square}{N_2}$.

We shall now show that the proposition holds true if $N_1 \leq 6$ or $N_2 \leq 6$. Suppose that $N_1 \leq 6$. Through use of the triangle inequality, we can show that

$$\|w - R_w^\alpha\|_\square = \left| \iint_{S\times T} (w - R_w^\alpha) dxdy \right| \leq \|(w - R_w^\alpha)\mathbb{1}_{S\times T}\|_1 \leq \|w\mathbb{1}_{S\times T}\|_1 + \|R_w^\alpha \mathbb{1}_{S\times T}\|_1.$$

Then, using Hölder's inequality (with $1/p + 1/q = 1$) and the fact that $|S| \leq 6\beta$, we can show

$$\|w\mathbb{1}_{S\times T}\|_1 + \|R_w^\alpha \mathbb{1}_{S\times T}\|_1 \leq \|w\|_p \|\mathbb{1}_{S\times T}\|_q + \|R_w^\alpha\|_\infty \|\mathbb{1}_{S\times T}\|_1 \leq (6\beta)^{1-\frac{1}{p}} + \alpha^{\frac{-2}{p}}(6\beta),$$

where the bound for $\|R_w^\alpha\|_\infty$ is given by Lemma 4.2.6 (ii). Furthermore, as this upper bound holds for all $\beta > \alpha$, it must be true that $\|w - R_w^\alpha\|_\square \leq (6\alpha)^{1-\frac{1}{p}} + \alpha^{\frac{-2}{p}}(6\alpha) \leq 12\alpha^{1-\frac{2}{p}}$. Thus, remembering that $\alpha = \|w\|_\infty^{-\frac{p}{3p-2}} \Lambda(w)^{\frac{2p}{5p-2}}$, we can say that

$$\|w - R_w^\alpha\|_\square \leq 12\alpha^{1-\frac{2}{p}} = 12\|w\|_\infty^{-\frac{p-2}{3p-2}} \Lambda(w)^{\frac{2p-4}{5p-2}}.$$

Furthermore, as $\|w\|_\infty \geq 1$, $\Lambda(w) \leq 1$, and $p > 2$, we get that

$$\begin{aligned} 12\|w\|_\infty^{-\frac{p-2}{3p-2}} \Lambda(w)^{\frac{2p-4}{5p-2}} &\leq 12\|w\|_\infty^{\frac{2p}{3p-2}} \Lambda(w)^{\frac{p-2}{5p-2}} \\ &\leq 6\left(1 + 2\|w\|_\infty^{\frac{2p}{3p-2}}\right)\Lambda(w)^{\frac{p-2}{5p-2}} \\ &\leq 6\left(1 + (8\|w - R_w^\alpha\|_p + 2)\|w\|_\infty^{\frac{2p}{3p-2}}\right)\Lambda(w)^{\frac{p-2}{5p-2}}, \end{aligned}$$

proving the statement of the proposition. Thus, we can assume that $N_1 \geq 7$, and by an identical argument we can assume the same for N_2.

Now, with repeated applications of the triangle inequality, we get that

$$\sum_{\substack{1 \leq i < N_1,\ 1 \leq j < N_2 \\ (S_i \times T_j) \cap \Delta \neq \emptyset}} \left| \iint_{S_i \times T_j} (w - R_w^\alpha) dxdy \right| \geq \left| \sum_{\substack{1 \leq i < N_1,\ 1 \leq j < N_2 \\ (S_i \times T_j) \cap \Delta \neq \emptyset}} \iint_{S_i \times T_j} (w - R_w^\alpha) dxdy \right|$$

$$
\begin{aligned}
&\geq \left| \iint_{(S\times T)\cap\Delta} (w - R_w^\alpha)dxdy \right| \\
&- \left| \iint_{(S_{N_1}\times T)\cup(S\times T_{N_2})} (w - R_w^\alpha)dxdy \right| \\
&\geq \left(\frac{1}{2} - \frac{1}{N_1} - \frac{1}{N_2}\right)\|R_w^\alpha - w\| \ . \qquad (4.17)
\end{aligned}
$$

Recall that by definition, $\Delta = \bigcup_{k=0}^{mM-1}\mathcal{R}_k \cup \bigcup_{k=1}^{mM-1}\mathcal{G}_k$, and that each of the regions $\mathcal{G}_k$ or $\mathcal{R}_k$ is bounded by boundary curves from the collection $\{f_k, g_l :\ 1 \leq k \leq mM-1, 1 \leq l \leq mM\}$ as defined in Remark 4.2.10. Thus, if a cell $S_i \times T_j$ does not cross the graph of any of these boundary curves, then it must be entirely contained inside one closed region $\overline{\mathcal{R}_k}$ or $\overline{\mathcal{G}_k}$. Next, by Lemma 4.2.12, none of the grey regions $\overline{\mathcal{G}_k}$ can contain any of the cells $S_i \times T_j$ with $1 \leq i < N_1$ and $1 \leq j < N_2$. We note that this is because $|S_i| = |T_j| = \beta$ for $1 \leq i < N_1$ and $1 \leq j < N_2$. Thus, these cells must either lie in a single region $\overline{\mathcal{R}_k}$ or cross a boundary curve. Let $\mathcal{I}$ denote the collection of indices (i, j) with $i < N_1$ and $j < N_2$, for which the associated cells do not lie in a single region $\overline{\mathcal{R}_k}$. From the above discussion, we have

$$
\mathcal{I} = \Big\{(i,j) : 1 \leq i < N_1, 1 \leq j < N_2, \text{ and } \exists\, 1 \leq k \leq mM-1 \text{ s.t. } (S_i \times T_j) \text{ crosses } f_k \text{ or } g_k, g_{mM}\Big\}.
$$

By Remark 4.2.10, the lower and upper boundaries f_k, g_k are increasing curves. As a result, each f_k (similarly g_k) crosses at most $2/\beta$ cells from the grid (see [40, Theorem 3.2] for a proof). Thus, we have

$$
|\mathcal{I}| \leq \frac{2(2mM-1)}{\beta}.
$$

Since every cell indexed in $\mathcal{I}$ is of size β^2, using Hölder's inequality, we show

$$
\sum_{\substack{(i,j)\,\in\,\mathcal{I} \\ (S_i \times T_j)\cap\Delta \neq \emptyset}} \left| \iint_{S_i\times T_j} (w - R_w^\alpha)dxdy \right| \leq \frac{2(2mM-1)}{\beta}\|w - R_w^\alpha\|_p \beta^{2-\frac{2}{p}}. \qquad (4.18)
$$

By inequalities (4.17) and (4.18), we get

$$\sum_{\substack{1 \le i < N_1,\ 1 \le j < N_2 \\ (S_i \times T_j) \cap \Delta \ne \emptyset \\ (i,j) \notin \mathcal{I}}} \left| \iint_{S_i \times T_j} (w - R_w^\alpha) dxdy \right|$$
$$\ge \left(\frac{1}{2} - \frac{1}{N_1} - \frac{1}{N_2}\right) \|R_w^\alpha - w\| \quad - \frac{2(2mM-1)}{\beta} \|w - R_w^\alpha\|_p \beta^{2-\frac{2}{p}}.$$

By the pigeonhole principle, since there are at most $(\lfloor \beta^{-1} \rfloor)^2$ cells $S_i \times T_j$ of size $\beta \times \beta$ and $\beta^2 \le (\lfloor \beta^{-1} \rfloor)^{-2}$, there must exist a cell $S_{i_0} \times T_{j_0} \subseteq \Delta$ so that $|S_{i_0}| = |T_{j_0}| = \beta$, $(i_0, j_0) \notin \mathcal{I}$ and

$$\left| \iint_{S_{i_0} \times T_{j_0}} (w - R_w^\alpha) dxdy \right| \ge \beta^2 \left[\left(\frac{1}{2} - \frac{1}{N_1} - \frac{1}{N_2}\right) \|R_w^\alpha - w\| \ \right]$$
$$- \beta^2 \left[\frac{2(2mM-1)}{\beta} \|w - R_w^\alpha\|_p \beta^{2-\frac{2}{p}} \right].$$

So $S_{i_0} \times T_{j_0}$ lies entirely in $\overline{\mathcal{R}_k} = \overline{\mathcal{B}_k \cap \mathcal{W}_{k+1}}$ for some $0 \le k \le mM - 1$. By Lemma 4.2.14, we can reduce β to α; this lets us assume that $S_{i_0} \times T_{j_0}$ is an $\alpha \times \alpha$ cell contained in $\mathcal{R}_k = \mathcal{B}_k \cap \mathcal{W}_{k+1}$ satisfying

$$\left| \iint_{S_{i_0} \times T_{j_0}} (w - R_w^\alpha) dxdy \right| \ge \alpha^2 \left[\left(\frac{1}{2} - \frac{1}{N_1} - \frac{1}{N_2}\right) \|R_w^\alpha - w\| \ \right]$$
$$- \alpha^2 \left[\frac{2(2mM-1)}{\beta} \|w - R_w^\alpha\|_p \beta^{2-\frac{2}{p}} \right].$$

As this inequality holds for all $\beta > \alpha$, we take the limit as $\beta \to \alpha$ to get that

$$\left| \iint_{S_{i_0} \times T_{j_0}} (w - R_w^\alpha) dxdy \right| \ge \alpha^2 \left[\left(\frac{1}{2} - \frac{1}{N_1} - \frac{1}{N_2}\right) \|R_w^\alpha - w\| \ \right]$$
$$- \alpha^2 \left[2(2mM-1) \|w - R_w^\alpha\|_p \alpha^{1-\frac{2}{p}} \right]. \tag{4.19}$$

From the definition of R_w^α, we observe that if $1 \le k \le Mm - 1$, then we have $\frac{k}{m} < R_w^\alpha \le \frac{k+1}{m}$ on $S_{i_0} \times T_{j_0}$; if $k = 0$, then $0 \le R_w^\alpha \le \frac{1}{m}$ on $S_{i_0} \times T_{j_0}$.

Claim 4.2.16. *Under the assumptions made so far,*

$$\|R_w^\alpha - w\|_\square \le 6\left(\frac{2\Lambda(w)}{\alpha^2} + \frac{1}{m} + 2(2mM-1)\|w - R_w^\alpha\|_p \alpha^{1-\frac{2}{p}}\right). \tag{4.20}$$

To prove the claim, we consider three cases:

Case 1: Assume that $\iint_{S_{i_0}\times T_{j_0}}(w - R_w^\alpha)dxdy > 0$ and $0 \le k \le mM-2$. In this case, using (4.19), we have

$$\begin{aligned}\overline{w}(S_{i_0}\times T_{j_0}) - \frac{k}{m} &\ge \overline{w - R_w^\alpha}(S_{i_0}\times T_{j_0})\\ &\ge \left(\frac{1}{2} - \frac{1}{N_1} - \frac{1}{N_2}\right)\|R_w^\alpha - w\|_\square - 2(2mM-1)\|w - R_w^\alpha\|_p \alpha^{1-\frac{2}{p}},\end{aligned} \tag{4.21}$$

where we use that $|S_{i_0}\times T_{j_0}| = \alpha^2$.

Now let (x, y) be the lower right corner of $S_{i_0}\times T_{j_0}$. Then $(x, y) \in \mathcal{W}_{k+1}$, implying $\text{LR}(x, y)$ contains a region $S_l \times T_l$ so that $|S_l| = |T_l| = \alpha$, and $\overline{w}(S_l\times T_l) \le \frac{k+1}{m}$. So inequality (4.21) combined with Remark 4.2.4 implies that

$$\Lambda(w) \ge \frac{\alpha^2}{2}\left(\left(\frac{1}{2} - \frac{1}{N_1} - \frac{1}{N_2}\right)\|R_w^\alpha - w\|_\square - 2(2mM-1)\|w - R_w^\alpha\|_p\alpha^{1-\frac{2}{p}} + \frac{k}{m} - \frac{k+1}{m}\right).$$

We conclude that

$$\begin{aligned}\frac{1}{6}\|R_w^\alpha - w\|_\square &\le \left(\frac{1}{2} - \frac{1}{N_1} - \frac{1}{N_2}\right)\|R_w^\alpha - w\|_\square\\ &\le \frac{2\Lambda(w)}{\alpha^2} + \frac{1}{m} + 2(2mM-1)\|w - R_w^\alpha\|_p\alpha^{1-\frac{2}{p}}\\ &\le \frac{2\Lambda(w)}{\alpha^2} + \frac{1}{m} + 4mM\|w - R_w^\alpha\|_p\alpha^{1-\frac{2}{p}},\end{aligned}$$

where in the first inequality, we used the fact that both N_1 and N_2 are at least 7.

Case 2: For the second case, assume $\iint_{S_{i_0}\times T_{j_0}}(R_w^\alpha - w)dxdy \ge 0$ and $1 \le k \le$

$mM-1$. By a similar argument used to show (4.21),

$$\frac{k+1}{m} - \overline{w}(S_{i_0} \times T_{j_0}) \geq \left(\frac{1}{2} - \frac{1}{N_1} - \frac{1}{N_2}\right)\|R_w^\alpha - w\| \ \ - 2(2mM-1)\|w - R_w^\alpha\|_p \alpha^{1-\frac{2}{p}}.$$

Now let (x,y) be the upper left corner of $S_{i_0} \times T_{j_0}$. Then $(x,y) \in \mathcal{B}_k$, which means $\mathrm{UL}(x,y)$ contains a region $U \times V$ such that $|U| = |V| = \alpha$ and $\overline{w}(U \times V) > \frac{k}{m}$. Using the definition of Λ, similar to the argument in Case 1, we get

$$\Lambda(w) \ \ \geq \ \ \frac{\alpha^2}{2}\left(\frac{k}{m} + \left(\frac{1}{2} - \frac{1}{N_1} - \frac{1}{N_2}\right)\|R_w^\alpha - w\| \ \ - 2(2mM-1)\|w - R_w^\alpha\|_p \alpha^{1-\frac{2}{p}} - \frac{k+1}{m}\right).$$

Once more assuming that both N_1 and N_2 are at least 7, we get

$$\frac{1}{6}\|R_w^\alpha - w\| \ \ \leq \frac{2\Lambda(w)}{\alpha^2} + \frac{1}{m} + 2(2mM-1)\|w - R_w^\alpha\|_p \alpha^{1-\frac{2}{p}}.$$

Case 3: Assume that either $\iint_{S_{i_0} \times T_{j_0}} (w - R_w^\alpha) dxdy > 0$ and $k = mM-1$ *or* that $\iint_{S_{i_0} \times T_{j_0}} (R_w^\alpha - w) dxdy \geq 0$ and $k = 0$. We note that $\mathcal{R}_{mM-1} = \mathcal{B}_{mM-1} \cap \mathcal{W}_{mM} = \mathcal{B}_{mM-1}$ and that $\mathcal{R}_0 = \mathcal{B}_0 \cap \mathcal{W}_1 = \mathcal{W}_1$. In the first assumption, we have that $R_w^\alpha > M - \frac{1}{m}$ on $S_{i_0} \times T_{j_0}$ whereas $w \leq M$ by definition. Thus, $\frac{\alpha^2}{m} \geq \iint_{S_{i_0} \times T_{j_0}} (w - R_w^\alpha) dxdy > 0$. In the second assumption, we have that $R_w^\alpha \leq \frac{1}{m}$ on $S_{i_0} \times T_{j_0}$ and by positivity of the integral in the second assumption it must be that $\frac{\alpha^2}{m} \geq \iint_{S_{i_0} \times T_{j_0}} R_w^\alpha - w \geq 0$. Combining either of these results with (4.19) gives us that

$$\frac{\alpha^2}{m} \geq \alpha^2 \left[\left(\frac{1}{2} - \frac{1}{N_1} - \frac{1}{N_2}\right)\|R_w^\alpha - w\| \ \ - 2(2mM-1)\|w - R_w^\alpha\|_p \alpha^{1-\frac{2}{p}}\right].$$

This can be rearranged (once again using that $N_1, N_2 \geq 7$) to show

$$\frac{1}{6}\|R_w^\alpha - w\| \ \ \leq \frac{1}{m} + 2(2mM-1)\|w - R_w^\alpha\|_p \alpha^{1-\frac{2}{p}} \leq \frac{1}{m} + 2(2mM-1)\|w - R_w^\alpha\|_p \alpha^{1-\frac{2}{p}} + \frac{2\Lambda(w)}{\alpha^2},$$

which is our desired result.

So inequality (4.20) holds in all cases. Now taking $\alpha = \|w\|_\infty^{-\frac{p}{3p-2}} \Lambda(w)^{\frac{2p}{5p-2}}$ and

$m = \lceil \Lambda(w)^{-\frac{p-2}{5p-2}} \rceil$ in (4.20), we get

$$\frac{1}{6}\|R_w^\alpha - w\| \leq \left(1+(4\|w-R_w^\alpha\|_p+2)\|w\|_\infty^{\frac{2p}{3p-2}}\right)\Lambda(w)^{\frac{p-2}{5p-2}} + \left(4\|w-R_w^\alpha\|_p\|w\|_\infty^{\frac{2p}{3p-2}}\right)\Lambda(w)^{\frac{2p-4}{5p-2}}.$$

Since $\Lambda(w) \leq 1$, we get $\Lambda(w)^{\frac{2p-4}{5p-2}} \leq \Lambda(w)^{\frac{p-2}{5p-2}}$; this simplifies the above equation to the desired result. □

Proposition 4.2.15 can also be used to make statements about traditional graphons and kernels.

Corollary 4.2.17. *Suppose $w : [0,1]^2 \to [0,1]$ is a graphon. Then*

$$\|R_w^\alpha - w\| \leq 66\Lambda(w)^{\frac{1}{5}},$$

where $\alpha = \|w\|_\infty^{-\frac{1}{3}}\Lambda(w)^{\frac{2}{5}}$. For a bounded symmetric function $u : [0,1]^2 \to \mathsf{R}$ we have

$$\|R_u^{\alpha^*} - u\| \leq 66\|u\|_\infty^{\frac{4}{5}}\Lambda(u)^{\frac{1}{5}},$$

where $\alpha^* = \|u\|_\infty^{-\frac{2}{5}}\Lambda(u)^{\frac{2}{5}}$.

Proof. From (4.15), we have $\|R_w^{\alpha^p} - w\| \leq 66\Lambda(w)^{\frac{p-2}{5p-2}}$, since $\|w\|_\infty \leq 1$ and $\|w - R_w^{\alpha^p}\|_p \leq 1$. Here, $\alpha^p = \|w\|_\infty^{-\frac{p}{3p-2}}\Lambda(w)^{\frac{2p}{5p-2}}$ as (4.15) is dependent upon the parameter chosen for R_w^α which varies dependent on p. However, as $L^\infty[0,1]^2 \subseteq L^p[0,1]^2$ for every $p > 1$, we can allow $p \to \infty$, showing $\|R_w^\alpha - w\| \leq 66\Lambda(w)^{\frac{1}{5}}$, where $\alpha = \|w\|_\infty^{-\frac{1}{3}}\Lambda(w)^{\frac{2}{5}}$ as above.

For a kernel $u : [0,1]^2 \to [0, \|u\|_\infty]$, we scale by the infinity norm to make a new function $u^* := u/\|u\|_\infty$ to get that

$$\|R_{u^*}^{\alpha^*} - u^*\| \leq 66\Lambda(u^*)^{\frac{1}{5}}, \tag{4.22}$$

where $\alpha^* = \|u\|_\infty^{-\frac{2}{5}} \Lambda(u)^{\frac{2}{5}}$. We note that by definition of Λ and positivity of $\|u\|_\infty$, it must be the case that $\Lambda(u^*)^{\frac{1}{5}} = \|u\|_\infty^{-\frac{1}{5}} \Lambda(u)^{\frac{1}{5}}$. Furthermore, by definition of R_u^α, we have

$$R_{u^*}^{\alpha^*} = \|u\|_\infty^{-1} R_u^{\alpha^*}.$$

Thus, combining these two observations with (4.22) yields

$$\frac{1}{\|u\|_\infty} \|R_u^{\alpha^*} - u\| \quad \leq 66 \|u\|_\infty^{-\frac{1}{5}} \Lambda(u)^{\frac{1}{5}},$$

which implies that

$$\|R_u^{\alpha^*} - u\| \quad \leq 66 \|u\|_\infty^{\frac{4}{5}} \Lambda(u)^{\frac{1}{5}},$$

proving the claim. □

Now that we have proven both Proposition 4.2.15 and a useful corollary, we are ready to state and prove our main result about stability of Λ. We begin with a necessary definition—a key technique in this proof is taking an unbounded graphon and "cutting it off" at a certain threshold. We introduce notation for that basic concept below.

Definition 4.2.18 (M-cut-off)**.** Let $w \in \mathcal{W}^1$, let

$$E_M = \{(x, y) \in [0, 1]^2 : w(x, y) > M\},$$

let $\mathbb{1}_M$ be the characteristic function of E_M, and let $\mathbb{1}$ be the characteristic function of $[0, 1]^2$. The M-*cut-off* of w, denoted by w_M, is defined to be $w_M := w(\mathbb{1} - \mathbb{1}_M)$.

Theorem 4.2.19. *Suppose* $w : [0, 1]^2 \to [0, \infty)$ *is an* L^p*-kernel with* $p > 5$ *and* $\|w\|_p = 1$. *Then there exists some* $\alpha \in [0, \frac{1}{2})$ *such that* R_w^α*, the Robinson approximation of* w *with parameter* α*, satisfies*

$$\|R_w^\alpha - w\| \quad \leq 119 \Lambda(w)^{\frac{p-5}{5p-5}}.$$

The idea of the proof is as follows: We use an $\epsilon/3$ argument on the difference in cut norm between w and its Robinson approximation R^α_w, introducing the terms w_M and $R^\alpha_{w_M}$, bounding each term above in terms of Λ. If w_M is Robinson, then the upper bound is proved only using properties of Robinson functions without use of Corollary 4.2.17. If w_M is not Robinson, then its difference in cut norm with $R^\alpha_{w_M}$ must be handled using Corollary 4.2.17. Handling these two cases finishes the proof.

Proof. By definition of Robinson approximation, if $\Lambda(w) = 0$, then we set $\alpha = 0$, resulting in $R^\alpha_w = w$. Thus, we assume that $\Lambda(w) > 0$ and define our cut-off value $M = 2\Lambda(w)^{-\frac{1}{p-1}}$. The proof then breaks into two cases.

Case 1: Suppose that $\Lambda(w_M) > 0$. We begin by defining the Robinson parameter $\alpha = \|w_M\|_\infty^{-\frac{2}{5}}\Lambda(w_M)^{\frac{2}{5}}$, where w_M is as defined in Definition 4.2.18. The triangle inequality can then be used to show that

$$\|R^\alpha_w - w\| \leq \|w - w_M\| + \|R^\alpha_{w_M} - w_M\| + \|R^\alpha_w - R^\alpha_{w_M}\| .$$

We will proceed by dealing with each of these terms one by one, starting with $\|w - w_M\|$. Define $\mathbb{1}_M$ to be the characteristic function of E_M, the region of $[0,1]^2$ where $w > M$. Since $M|E_M|^{\frac{1}{p}} \leq \|w\mathbb{1}_M\|_p \leq \|w\|_p = 1$, we get that $|E_M| \leq (\frac{1}{M})^p$. Therefore, letting $1/p + 1/q = 1$, we get that

$$\|\mathbb{1}_M\|_q = |E_M|^{1/q} \leq \left(\frac{1}{M}\right)^{\frac{p}{q}} = M^{1-p} = 2^{1-p}\Lambda(w). \tag{4.23}$$

It is also true that

$$\|w - w_M\| \leq \|w - w_M\|_1 = \|w\mathbb{1}_M\|_1 \leq \|w\|_p\|\mathbb{1}_M\|_q \leq 2^{1-p}\Lambda(w) \leq 2^{1-p}\Lambda(w)^{\frac{p-5}{5p-5}}, \tag{4.24}$$

handling the first term. We can further say that $\|w - w_M\| \leq \Lambda(w)^{\frac{p-5}{5p-5}}$ for all $p > 5$. Now we shift focus to $\|R^\alpha_{w_M} - w_M\|$. By Corollary 4.2.17, as w_M is bounded, we have the following.

$$\begin{aligned} \|R^{\alpha}_{w_M} - w_M\| &\leq 66\|w_M\|_{\infty}^{\frac{4}{5}} \Lambda(w_M)^{\frac{1}{5}} \\ &\leq 66M^{\frac{4}{5}}(\Lambda(w) + 2^{1-p}\Lambda(w))^{\frac{1}{5}} \\ &\leq 66 \cdot (1 + 2^{1-p})^{\frac{1}{5}} M^{\frac{4}{5}} \Lambda(w)^{\frac{1}{5}} \\ &\leq 66 \cdot (1 + 2^{1-p})^{\frac{1}{5}} \cdot 2^{\frac{4}{5}} \Lambda(w)^{\frac{p-5}{5p-5}}, \end{aligned} \tag{4.25}$$

where the second inequality is due to the combination of Proposition 4.2.3 (iii) for $w = w_M + w\mathbb{1}_M$ alongside the fact that $\Lambda(w\mathbb{1}_M) \leq \|w\mathbb{1}_M\|_1 \leq 2^{1-p}\Lambda(w)$ by (4.23). We note that for $p > 5$, we have that $66 \cdot (1 + 2^{1-p})^{\frac{1}{5}} \cdot 2^{\frac{4}{5}} \leq 117$; thus, we can say that $\|R^{\alpha}_{w_M} - w_M\| \ \leq 117\Lambda(w)^{\frac{p-5}{5p-5}}$.

We turn to the final term $\|R^{\alpha}_{w} - R^{\alpha}_{w_M}\|$ by first bounding the expression $\|R^{\alpha}_{w-w_M}\|$. We claim that $\|R^{\alpha}_{w-w_M}\|_{\infty} \leq \Lambda(w)^{\frac{p-5}{5p-5}}$. To show this, we let $(x, y) \in \Delta$ and $\epsilon > 0$. Then, there must exist some measurable sets $S, T \subset [0, 1]$ such that $|S| = |T| = \alpha$ and

$$\begin{aligned} |R^{\alpha}_{w-w_M}(x, y)| - \epsilon &\leq \frac{1}{\alpha^2}\left| \iint_{S\times T} (w - w_M)dxdy \right| \leq \frac{1}{\alpha^2}\|(w - w_M)\mathbb{1}_{S\times T}\|_1 \\ &= \frac{1}{\alpha^2}\|w\mathbb{1}_M\mathbb{1}_{S\times T}\|_1 \leq \frac{1}{\alpha^2}\|w\|_p\|\mathbb{1}_M\|_{r_1}\|\mathbb{1}_{S\times T}\|_{r_2} \\ &\leq M^{-\frac{p}{r_1}}\alpha^{\frac{2}{r_2}-2} = M^{-\frac{p}{r_1}}\alpha^{-\frac{2}{p}-\frac{2}{r_1}}, \end{aligned} \tag{4.26}$$

where $\frac{1}{p} + \frac{1}{r_1} + \frac{1}{r_2} = 1$. Letting $r_1 = \frac{p}{p-1}$ gets us

$$\begin{aligned} |R^{\alpha}_{w-w_M}(x, y)| - \epsilon \leq M^{-\frac{p}{r_1}}\alpha^{-\frac{2}{p}-\frac{2}{r_1}} &= M^{-(p-1)}\alpha^{-\frac{2}{p}-\frac{2(p-1)}{p}} = 2^{1-p}\Lambda(w)\alpha^{-2} \\ &= 2^{1-p}\Lambda(w)(\|w_M\|_{\infty}^{-\frac{2}{5}}\Lambda(w_M)^{\frac{2}{5}})^{-2} \\ &= 2^{1-p}\Lambda(w)\|w_M\|_{\infty}^{\frac{4}{5}}\Lambda(w_M)^{-\frac{4}{5}} \\ &\leq 2^{1-p}\Lambda(w)\|w_M\|_{\infty}^{\frac{4}{5}}(|\Lambda(w) - M^{1-p}|)^{-\frac{4}{5}} \\ &\leq 2^{1-p}\Lambda(w)M^{\frac{4}{5}}(1 - 2^{1-p})^{-\frac{4}{5}}\Lambda(w)^{-\frac{4}{5}} \end{aligned}$$

$$\leq \frac{2^{\frac{9-5p}{5}}}{(1-2^{1-p})^{\frac{4}{5}}} \Lambda(w)^{\frac{p-5}{5p-5}},$$

where the second inequality is due to the combination of (4.23) and the following:

$$\Lambda(w_M) = \Lambda(w(\mathbb{1} - \mathbb{1}_M)) \geq \Lambda(w) - \Lambda(w\mathbb{1}_M) \geq \Lambda(w) - \|w\mathbb{1}_M\|_1 \geq (1-2^{1-p})\Lambda(w) > 0.$$

Letting ϵ go to 0 alongside noting that $2^{\frac{9-5p}{5}}(1-2^{1-p})^{-\frac{4}{5}} \leq 1$ for $p > 5$ proves the claim. By Lemma 4.2.6 (i), $0 \leq R^\alpha_w(x,y) - R^\alpha_{w_M}(x,y) \leq R^\alpha_{w-w_M}(x,y)$, so

$$\|R^\alpha_w - R^\alpha_{w_M}\| \leq \|R^\alpha_w - R^\alpha_{w_M}\|_\infty \leq \|R^\alpha_{w-w_M}\|_\infty \leq \Lambda(w)^{\frac{p-5}{5p-5}}. \tag{4.27}$$

Thus (4.24), (4.25), and (4.27) together will give us that $\|R^\alpha_w - w\| \leq 119\Lambda(w)^{\frac{p-5}{5p-5}}$, proving the statement of the theorem.

Case 2: Suppose that $\Lambda(w_M) = 0$. We now define our Robinson parameter $\alpha = M^{-\frac{2}{5}}\Lambda(w)^{\frac{2}{5}}$. We proceed similarly to Case 1, using the triangle inequality to show that

$$\|R^\alpha_w - w\| \leq \|w - w_M\| + \|R^\alpha_{w_M} - w_M\| + \|R^\alpha_w - R^\alpha_{w_M}\| .$$

The first and third term on the right side of the inequality can be handled identically to Case 1, yielding $\|w - w_M\| \leq \Lambda(w)^{\frac{p-5}{5p-5}}$ and $\|R^\alpha_w - R^\alpha_{w_M}\| \leq \Lambda(w)^{\frac{p-5}{5p-5}}$ when the new value of α is used. However, the second term $\|R^\alpha_{w_M} - w_M\|$ must be handled differently. We observe that as $\Lambda(w_M) = 0$, it must be that w_M is Robinson a.e.; we therefore can directly approximate $R^\alpha_{w_M}$ as in the following claim.

Claim 4.2.20. *When $w_M : [0,1]^2 \to [0,M]$ is Robinson a.e., we have*

$$\|w_M - R^\alpha_{w_M}\| \leq \iint_{\{|x-y|\leq 2\alpha\}} w_M(x,y)dxdy.$$

To prove the claim, first note that for any point $(x,y) \in \Delta$, we have that $w_M(x,y)$ is an upper bound for every value of w_M over the set UL(x,y); thus, any average over that set (such as in the definition of R^α_w) would not exceed $w_M(x,y)$. Since w_M is non-negative, this observation implies that $R^\alpha_{w_M} \leq w_M$. On the other hand, note that as w_M is Robinson a.e., we have

$$R^\alpha_{w_M}(x,y) = \begin{cases} \dfrac{1}{\alpha^2}\displaystyle\iint_{[x-\alpha,x]\times[y,y+\alpha]} w_M dxdy & \text{for } (x,y)\in[\alpha,1-\alpha]^2\cap\Delta \\ 0 & \text{otherwise} \end{cases}, \tag{4.28}$$

Where we also define $R^\alpha_{w_M}(x,y) = R^\alpha_{w_M}(y,x)$. We now introduce an auxiliary function $\widetilde{w}_M$ defined as follows:

$$\widetilde{w}_M(x,y) = \widetilde{w}_M(y,x) = \begin{cases} w_M(x-\alpha, y+\alpha) & (x,y)\in[\alpha,1-\alpha]^2\cap\Delta \\ 0 & \text{otherwise} \end{cases}.$$

It is clear from the definition of both $\widetilde{w}_M$ and (4.28) that $R^\alpha_{w_M} \geq \widetilde{w}_M$. Thus we have $0 \leq w_M - R^\alpha_{w_M} \leq w_M - \widetilde{w}_M$ pointwise, allowing us to show the following:

$$\begin{aligned} \|w_M - R^\alpha_{w_M}\| &\leq \|w_M - \widetilde{w}_M\| = 2\iint_{[0,1]^2\cap\Delta}(w_M - \widetilde{w}_M)dxdy \\ &= 2\left(\iint_{[0,1]^2\cap\Delta} w_M dxdy - \iint_{[\alpha,1-\alpha]^2\cap\Delta}\widetilde{w}_M dxdy\right) \\ &= 2\left(\iint_{[0,1]^2\cap\Delta} w_M dxdy - \iint_{0\leq x\leq y-2\alpha\leq 1-2\alpha} w_M(x,y)dxdy\right) \\ &\leq \iint_{\{|x-y|\leq 2\alpha\}} w_M(x,y)dxdy. \end{aligned}$$

This proves Claim 4.2.20. Finally, note that

$$\begin{aligned} \iint_{\{|x-y|\leq 2\alpha\}} w_M(x,y)dxdy &\leq \|w_M\|_p \|\mathbb{1}_{[0,1]^2\cap\{|x-y|\leq 2\alpha\}}\|_q \leq (1-(1-2\alpha)^2)^{\frac{1}{q}} \\ &\leq 4\alpha^{1-\frac{1}{p}} = 4\Big(\frac{\Lambda(w)}{M}\Big)^{(\frac{2}{5})(1-\frac{1}{p})} \leq 4\Lambda(w)^{\frac{2}{5}} \leq 4\Lambda(w)^{\frac{p-5}{5p-5}}. \end{aligned}$$

This last inequality holds because $\Lambda(w) \leq 1$ and $\frac{2}{5} \geq (p-5)(5p-5)^{-1}$ for all $p > 5$. Therefore, if $\Lambda(w_M) = 0$, we have that $\|w - R^\alpha_w\| \leq 6\Lambda(w)^{\frac{p-5}{5p-5}}$, implying the original claim. Thus the theorem holds true. □

4.2.3 Approximating in cut norm

In this section we use the tools we have developed so far to study how Λ interacts with sequences of sparse graphs; we shall show that it respects sparse graph convergence. However, unlike working with L^p graphons, for a sequence of sparse graphs $\{G_n\}$ it is the convergence of $\{G_n/\|G_n\|_1\}$ in δ that needs to be considered. This presents an issue, as if w is a graphon and $\{G_n\}$ is a sparse graph sequence such that $\delta\ (G_n/\|G_n\|_1, w) \to 0$, we can apply an ordering ϕ_n to G_n such that if $w_n := w_{G_n^{\phi_n}}$, then $\|w_n/\|w_n\|_1 - w\| \to 0$. Thus, we can show that

$$\Lambda(w_n) = \Lambda\left(\frac{w_n}{\|w_n\|_1} \cdot \|w_n\|_1\right) = \|w_n\|_1 \Lambda\left(\frac{w_n}{\|w_n\|_1}\right) \to 0 \cdot \Lambda(w) = 0.$$

So we can not simply consider Λ applied to the associated graphons of a sparse graph sequence. With this in mind, we introduce a slightly different parameter that handles this issue.

Definition 4.2.21 (Robinson parameter for sparse graph sequences)**.** Let $w \in \mathcal{W}^1$. Then we define the parameter Λ^* as follows:

$$\Lambda^*(w) := \Lambda\left(\frac{w}{\|w\|_1}\right).$$

If G is a labelled graph, we define $\Lambda^*(G) = \Lambda^*(w_G)$, where w_G is the associated graphon of G with respect to its labelling.

Combining this definition with the results of the previous section, we are ready to present our main result.

Theorem 4.2.22. *Let $\{G_n\}$ be a sequence of simple labelled graphs such that $\delta\ (G_n/\|G_n\|_1, w) \to 0$. Then there exists an ordering ϕ_n of the vertices of G_n such that*

$$\Lambda^*(G_n^{\phi_n}) \to \Lambda^*(w)$$

as $n \to \infty$.

Proof. Let $\{G_n\}$ be a sequence of simple labelled graphs such that $\delta_\square(G_n/\|G_n\|_1, w) \to 0$. This implies that there is an ordering ϕ_n of the vertices of G_n such that $\|w_{G_n^{\phi_n}}/\|w_{G_n}\|_1 - w\|_\square \to 0$. For ease of readability, we shall let $w_n := w_{G_n^{\phi_n}}$. In order to show that this ordering ϕ_n satisfies the claim of the theorem, we shall first prove that $\|w\|_1 = 1$. We note that as w_n is a nonnegative function, $\|w_n\|_\square = \|w_n\|_1$. Combining this knowledge with the reverse triangle inequality, it must be the case that

$$\begin{aligned} \left| \left\| \frac{w_n}{\|w_n\|_1} \right\|_\square - \|w\|_\square \right| &\leq \left\| \frac{w_n}{\|w_n\|_1} - w \right\|_\square \\ \left|1 - \|w\|_\square \right| &\leq \left\| \frac{w_n}{\|w_n\|_1} - w \right\|_\square \end{aligned}$$

which implies that $|1 - \|w\|_\square| = 0$ as the above inequality holds for all $n \in \mathbb{N}$. Thus, $\|w\|_\square = 1$, but because w is a nonnegative function, $\|w\|_1 = \|w\|_\square = 1$ as well. Therefore, by Proposition 4.2.3 (i),

$$\begin{aligned} |\Lambda^*(w_n) - \Lambda^*(w)| &= \left| \Lambda\left(\frac{w_n}{\|w_n\|_1}\right) - \Lambda\left(\frac{w}{\|w\|_1}\right) \right| \\ &\leq 2 \left\| \frac{w_n}{\|w_n\|_1} - \frac{w}{\|w\|_1} \right\|_\square = 2 \left\| \frac{w_n}{\|w_n\|_1} - w \right\|_\square \to 0, \end{aligned}$$

proving the initial claim. □

4.2.4 Proofs of Lemmas 4.2.12, 4.2.13, and 4.2.14

In this subsection we present full proofs of the lemmas used previously in the chapter, recalling their statements for clarity.

Lemma. *(Lemma 4.2.12) Let $k \in \mathbb{Z}^{\geq 0}$, $w \in \mathcal{W}^p$, and $\alpha \in (0,1)$. Then, $\overline{\mathcal{G}_k}$ does not contain any $\beta \times \beta$ square, where $\beta > \alpha$. Here, $\overline{\mathcal{G}_k}$ denotes the closure of $\mathcal{G}_k$ in the Euclidean topology of $\mathbb{R}^2$.*

Proof of Lemma 4.2.12. Let $k \in \mathbb{Z}^{\geq 0}$ be fixed and let $\mathcal{G}_k$ be defined with parameter m. By definition, points of $\overline{\mathcal{G}_k}$ lie on or between the lower and upper boundary curves of $\mathcal{G}_k$. Towards a contradiction, let $\beta > \alpha$ and suppose there exist measurable subsets $S, T \subseteq$

$[0,1]$ with $|S| = |T| = \beta$ for which $S \times T \subseteq \overline{\mathcal{G}_k}$. Let $a_1 = \inf S$, $a_2 = \sup S$, $b_1 = \inf T$, and $b_2 = \sup T$; note that $a_2 - a_1 \geq \beta$ and that $b_2 - b_1 \geq \beta$. Since $\overline{\mathcal{G}_k}$ is closed, we have that $(a_i, b_j) \in \mathcal{G}_k$ for $i, j = 1, 2$; we shall now show that $[a_1, a_2] \times [b_1, b_2] \subseteq \overline{\mathcal{G}_k}$. To do so, we suppose that some point $(z, w) \in \mathrm{UL}(a_2, b_1) \cap \mathrm{LR}(a_1, b_2) \setminus \mathcal{G}_k$, which implies that either $(z, w) \in \mathrm{UL}(a_2, b_1) \cap \mathrm{LR}(a_1, b_2) \cap \mathcal{B}_k$ or that $(z, w) \in \mathrm{UL}(a_2, b_1) \cap \mathrm{LR}(a_1, b_2) \cap \mathcal{W}_k$. Assuming neither of these sets are empty, the first case implies that $(a_2, b_1) \in \mathcal{B}_k$ while the second case implies that $(a_1, b_2) \in \mathcal{W}_k$, both of which are contradictions. Thus, $\mathrm{UL}(a_2, b_1) \cap \mathrm{LR}(a_1, b_2) \subseteq \mathcal{G}_k$ and so therefore $[a_1, a_2] \times [b_1, b_2] \subseteq \overline{\mathcal{G}_k}$ as desired.

Note that every point in $\overline{\mathcal{G}_k}$ that is not on the lower or upper boundary curves must be an interior point. Because both the lower and upper boundary functions of $\mathcal{G}_k$ are increasing, it must be that $(a_1, a_2) \times (b_1, b_2) \subseteq \mathcal{G}_k$. Clearly, $(a_1, a_2) \times (b_1, b_2)$ contains a closed $\alpha \times \alpha$ rectangle which we denote $[a_1', a_2'] \times [b_1', b_2']$. The two points (a_1', b_2') and (a_2', b_1') are elements of $\mathcal{G}_k$, so they fail to satisfy the conditions for both $\mathcal{W}_k$ and $\mathcal{B}_k$. In particular, we have $\overline{w}([a_1', a_2'] \times [b_1', b_2']) \leq (k-1)/m$ as $(a_2', b_1') \notin \mathcal{B}_k$, as well as $\overline{w}([a_1', a_2'], [b_1', b_2']) > (k-1)/m$ as $(a_1', b_2') \notin \mathcal{W}_k$. These statements form a contradiction, showing the initial claim must be true. $\square$

Lemma. *(Lemma 4.2.13) Let $u \in L^\infty([0,1])$, and $P \subseteq [0,1]$ be a measurable subset such that $\int_P u \neq 0$. Let $0 < \beta < |P|$ be xed. Then P can be partitioned into $N := \lceil |P|/\beta \rceil$ subsets $P_1, \ldots, P_N$ so that the following conditions are satis ed:*

(i) $P_1 \leq \ldots \leq P_{N-1}$.

(ii) $|P_i| = \beta$ *for* $1 \leq i \leq N-1$ *and* $|P_N| \leq \beta$.

(iii) $\left| \int_{P_N} u \right| \leq \frac{1}{N} \left| \int_P u \right|$.

Proof of Lemma 4.2.13. We begin by noting that for sets $P_1, \ldots, P_N$ as listed above that $|P_N| = \delta := |P| - \beta(\lceil \frac{|P|}{\beta} \rceil - 1)$. If $\delta = \beta$, by the pigeonhole principle we are done. Furthermore, given a subset of measure δ, one can create the other sets $P_1, \ldots, P_{N-1}$ by "filling in" $P \setminus P_N$ with consecutive sets of measure β. Thus, it is enough to find a subset of P of measure δ such that item (iii) is satisfied.

Without loss of generality, let $\int_P u > 0$. Therefore, there must exist some set $Q \subset P$ such that $|Q| = |P| - \delta(\lceil \frac{|P|}{\delta} \rceil - 1)$ and $\int_Q u > 0$. To see why, we can note that

$$\int_P u \; dxdy = \int_{P^+} u \; dxdy + \int_{P^-} u \; dxdy > 0, \tag{4.29}$$

where $P^+ := \{x \in P : u(x) > 0\}$ and $P^- := \{y \in P : u(y) < 0\}$. If $|P^+| \geq \delta$, Q can be taken as any subset of measure δ from P^+. If $|P^+| < \delta$, then we can take Q to be the union of P^+ and some subset of P^- whose total measure is δ; the integral of u over Q must be greater than 0 due to (4.29).

Towards a contradiction, assume that for any set S such that $|S| = \delta$, we have that

$$\left| \int_S u \; dxdy \right| > \frac{1}{N} \int_P u \; dxdy.$$

Consider now $R := P \setminus Q$. For any $x \in P$, let $r_x = \inf\{y : |R \cap [x, y]| = \delta\}$, and let the auxiliary function w be defined as

$$w(x) := \int_{R \cap [x, r_x]} u \; dxdy.$$

For ease of writing, let $R_x := R \cap [x, r_x]$. We claim that w is continuous; toward this end, let $\epsilon > 0$ and let $\delta = \|u\|_\infty^{-1}\epsilon$. Then, if $|x - y| < \delta$, we have that

$$\begin{aligned} |w(x) - w(y)| &= \left| \int_{R_x} u \; dxdy - \int_{R_y} u \; dxdy \right| \leq \left| \int_{R_x \Delta R_y} u \; dxdy \right| \\ &\leq \|u\|_\infty |R_x \Delta R_y| \leq \|u\|_\infty |x - y| < \epsilon, \end{aligned}$$

showing that w is indeed continuous. Obviously, $w(x)$ cannot achieve 0 at any x as otherwise R_x would a set of measure δ on which the integral of u is sufficiently small, violating our assumption. Suppose then there exist $x_0, y_0 \in R$ such that $w(x_0) > 0$ and $w(y_0) < 0$. Then, as w is continuous, there exists some z_0 in-between x_0 and y_0 such that $w(z_0) = 0$, again violating our assumption. Thus the only remaining possibility is that $w(x)$ is either strictly positive or strictly negative. Without loss of generality, we assume that $w(x) > 0$ for all x, and to avoid violating our assumption, it must also be that

$$w(x) > \frac{1}{N} \int_P u \; dxdy \tag{4.30}$$

for all $x \in R$. Consider $\{x_i\}_{i=1}^M \subset R$, where $x_1 = 0$, $x_{i+1} = r_{x_i}$ for $1 < i \leq M$, and $M = \lceil \frac{|P|}{\delta} \rceil - 1$. Then it must be the case, due to the positivity of w and (4.30), that

$$\int_P u \; dxdy = \sum_{i=1}^{M} w(x_i) + \int_Q u \; dxdy > \frac{M}{N} \int_P u \; dxdy. \tag{4.31}$$

We now claim that $M \geq N$, i.e. $\lceil \frac{|P|}{\delta} \rceil - 1 \geq \lceil \frac{|P|}{\beta} \rceil$. We note that as $\delta < \beta$, it must be the case that $M \geq N - 1$. Therefore, the only way our claim can be violated is if M is indeed equal to $N - 1$. For this to be possible, δ must be within a tight range of values: Specifically,

$$\frac{|P|}{\lceil \frac{|P|}{\beta} \rceil} \leq \delta < \beta.$$

However, the exact value of δ is known to be $|P| - \beta(\lceil \frac{|P|}{\beta} \rceil - 1)$. Thus, we get the following inequality:

$$\begin{aligned}
|P| - \beta\left(\left\lceil \frac{|P|}{\beta} \right\rceil - 1\right) &\geq \frac{|P|}{\lceil \frac{|P|}{\beta} \rceil} \\
|P| \left\lceil \frac{|P|}{\beta} \right\rceil - \beta \left\lceil \frac{|P|}{\beta} \right\rceil \left(\left\lceil \frac{|P|}{\beta} \right\rceil - 1\right) &\geq |P| \\
|P| \left\lceil \frac{|P|}{\beta} \right\rceil - |P| &\geq \beta \left\lceil \frac{|P|}{\beta} \right\rceil \left(\left\lceil \frac{|P|}{\beta} \right\rceil - 1\right) \\
|P| \left(\left\lceil \frac{|P|}{\beta} \right\rceil - 1\right) &\geq \beta \left\lceil \frac{|P|}{\beta} \right\rceil \left(\left\lceil \frac{|P|}{\beta} \right\rceil - 1\right) \\
|P| &\geq \beta \left\lceil \frac{|P|}{\beta} \right\rceil,
\end{aligned}$$

which occurs only when equality is achieved. However, that would then imply that $\delta = |P| - \beta(\lceil \frac{|P|}{\beta} \rceil - 1) = \beta$, which violates our assumptions and thus cannot happen.

Therefore, it must be the case that $M \geq N$, in which case (4.31) results in a contradiction. This implies the theorem must hold true and so such a partition of P must exist. $\square$

Lemma. *(Lemma 4.2.14) Let $f \in L^1([0,1]^2)$, and let $S, S' \subseteq [0,1]$ be measurable subsets such that $|S| = |S'|$. Suppose for a constant $C > 0$ we have*

$$\iint_{S\times S'} f \; dxdy \geq C.$$

Then, for every $\alpha \in (0,1)$, there exist measurable sets $T \subset S$ and $T' \subset S'$ such that $|T| = |T'| = \alpha|S|$ and

$$\frac{1}{|T\times T'|}\iint_{T\times T'} f \; dxdy \geq \frac{C}{|S\times S'|}.$$

Proof of Lemma 4.2.14. Suppose there exist integers n, k, l, with $l < k$, so that $|S| = \frac{k}{n}$ and $\alpha|S| = \frac{l}{n}$; the case where one or both of $|S|$ or α is not rational can be done using standard density/approximation arguments. Next, split S into k consecutive sets $S_1 \leq S_2 \leq \ldots \leq S_k$ of measure $\frac{1}{n}$; likewise, split S' into k consecutive sets $S'_1 \leq S'_2 \leq \ldots \leq S'_k$ also of measure $\frac{1}{n}$ and note that

$$\iint_{S\times S'} f \; dxdy = \sum_{i=1}^{k}\sum_{j=1}^{k}\iint_{S_i\times S'_j} f \; dxdy.$$

Let $B_{i,j} := \iint_{S_i\times S'_j} f$, and $\mathcal{N}_{k,l}$ denote the collection of all l-subsets of the set $\{1, ..., k\}$. Note that there are $\binom{k-1}{l-1}$ many l-subsets of k elements containing a specific element i_0. Using this, counting the number of times a specific $B_{i,j}$ appears in the following sum results in

$$\sum_{I\in\mathcal{N}_{k,l}}\sum_{J\in\mathcal{N}_{k,l}}\sum_{\substack{i\in I\\ j\in J}} B_{i,j} = \sum_{i,j=1}^{k}\binom{k-1}{l-1}^2 B_{i,j} \geq \binom{k-1}{l-1}^2 C.$$

Thus, by the pigeonhole principle, there exists set $I, J \in \mathcal{N}_{k,l}$ such that

$$\sum_{\substack{i \in I \\ j \in J}} B_{i,j} \geq \frac{\binom{k-1}{l-1}^2}{\binom{k}{l}^2} C = \frac{l^2}{k^2} C.$$

This implies that the sets $T = \cup_{i \in I} S_i$ and $T' = \cup_{i \in J} S_i'$ satisfy

$$\iint_{T \times T'} f \; dxdy \geq \frac{l^2}{k^2} C,$$

proving the initial statement. □

4.2.5 Open questions

This last subsection is dedicated to a collection of open questions that have appeared throughout work on this thesis.

Question 1: Does knowledge of the p-norm of a graphon w influence the p-norm of its corresponding Robinson approximation R_w^α? More specifically, if $\|w\|_p \leq 1$, is $\|R_w^\alpha\|_p \leq 1$?

Question 2: Is the value of R_w^α achieved by specific sets $A \times B$ at every point (x, y)?

INDEX

BIBLIOGRAPHY

[1] N. Alon and A. Naor. "Approximating the Cut-Norm via Grothendieck's Inequality". In: STOC '04. Chicago, IL, USA: Association for Computing Machinery, 2004, pp. 72–80. ISBN: 1581138520.

[2] J.R. Angelos, C.C. Cowen, and S.K. Narayan. "Triangular truncation and finding the norm of a Hadamard multiplier". In: *Linear Algebra and its Applications* 170 (1992), pp. 117–135. ISSN: 0024-3795.

[3] S. Armstrong, C. Guzmán, and C.A. Sing Long. "An Optimal Algorithm for Strict Circular Seriation". In: *SIAM Journal on Mathematics of Data Science* 3.4 (2021), pp. 1223–1250.

[4] J.E. Atkins, E.G. Boman, and B. Hendrickson. "A Spectral Algorithm for Seriation and the Consecutive Ones Problem". In: *SIAM Journal on Computing* 28.1 (1998), pp. 297–310.

[5] K. Azuma. "Weighted sums of certain dependent random variables". In: *Tohoku Math. J. (2)* 19.3 (1967), pp. 357–367.

[6] J.P. Barthélemy and F. Brucker. "NP-hard approximation problems in overlapping clustering". In: *Journal of classi cation* 18.2 (2001), pp. 159–183.

[7] G. Bennett. "Schur multipliers". In: *Duke Mathematical Journal* 44.3 (1977), pp. 603–639.

[8] G. Bennett. "Unconditional convergence and almost everywhere convergence". In: *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 34.2 (1976), pp. 135–155.

[9] Jacques Bertin. *Semiology of graphics*. Redlands, CA: ESRI Press, Jan. 2011.

[10] B. Bollobás, S. Janson, and O. Riordan. “The phase transition in inhomogeneous random graphs”. In: *Random Structures Algorithms* 31.1 (2007), pp. 3–122. ISSN: 1042-9832.

[11] B. Bollobás and O. Riordan. “Sparse graphs: Metrics and random models”. In: *Random Structures & Algorithms* 39.1 (2011), pp. 1–38.

[12] C. Borgs et al. “An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions”. In: *Transactions of the American Mathematical Society* 372.5 (May 2019), pp. 3019–3062. ISSN: 1088-6850.

[13] C. Borgs et al. “An L^p theory of sparse graph convergence II: LD convergence, quotients and right convergence”. In: *The Annals of Probability* 46.1 (Jan. 2018). ISSN: 0091-1798.

[14] G.W. Brainerd. “The Place of Chronological Ordering in Archaeological Analysis”. In: *American Antiquity* 16.4 (1951), pp. 301–313.

[15] X. Chang, D. Huang, and H. Wang. “A popularity-scaled latent space model for large-scale directed social network”. In: *Statist. Sinica* 29.3 (2019), pp. 1277–1299. ISSN: 1017-0405.

[16] Y. Cheng and G.M. Church. “Biclustering of Expression Data”. In: *Proceedings. International Conference on Intelligent Systems for Molecular Biology* 8 (2000), pp. 93–103.

[17] V. Chepoi, B. Fichet, and M. Seston. “Seriation in the Presence of Errors: NP-Hardness of l_∞-Fitting Robinson Structures to Dissimilarity Matrices”. In: *Journal of Classi cation* 26 (Dec. 2009), pp. 279–296.

[18] V. Chepoi and M. Seston. “Seriation in the Presence of Errors: A Factor 16 Approximation Algorithm for l_∞-Fitting Robinson Structures to Distances”. In: *Algorithmica* 59 (Feb. 2009), pp. 521–568.

[19] H. Chuangpishit et al. “Linear embeddings of graphs and graph limits”. In: *Journal of Combinatorial Theory, Series B* 113 (July 2015), pp. 162–184. ISSN: 0095-8956.

[20] C. Coine. “Schur multipliers on $\mathcal{B}(\mathcal{L}^{\checkmark}, \mathcal{L}^{\mathrm{II}})$”. In: *J. Operator Theory* 79.2 (2018), pp. 301–326. ISSN: 0379-4024.

[21] D.G. Corneil. “A simple 3-sweep LBFS algorithm for the recognition of unit interval graphs”. In: *Discrete Applied Mathematics* 138.3 (2004), pp. 371–379. ISSN: 0166-218X.

[22] D.G. Corneil et al. “Simple Linear Time Recognition of Unit Interval Graphs”. In: *Inf. Process. Lett.* 55 (1995), pp. 99–104.

[23] K.R. Davidson and A.P. Donsig. “Norms of Schur multipliers”. In: *Illinois J. Math.* 51.3 (2007), pp. 743–766. ISSN: 0019-2082.

[24] X. Deng, P. Hell, and J. Huang. “Linear-Time Representation Algorithms for Proper Circular-Arc Graphs and Proper Interval Graphs”. In: *SIAM Journal on Computing* 25.2 (1996), pp. 390–403.

[25] S.C. Dodd. “The Interrelation Matrix”. In: *Sociometry* 3.1 (1940), pp. 91–101. ISSN: 00380431. (Visited on 10/06/2022).

[26] P. G. Dodds et al. “Vilenkin systems and generalized triangular truncation operator”. In: *Integral Equations and Operator Theory* 40.4 (Dec. 2001), pp. 403–435. ISSN: 1420-8989.

[27] M. Dom. “Algorithimic Aspects of the Consecutive-Ones Property”. In: *Bull. EATCS* 98 (2009), pp. 27–59.

[28] R. G. Douglas. “Toeplitz Operators”. In: *Banach Algebra Techniques in Operator Theory*. New York, NY: Springer New York, 1998, pp. 158–184. ISBN: 978-1-4612-1656-8.

[29] R. Durrett. *Probability: Theory and Examples*. Thomson, 2005.

[30] X. Evangelopoulos et al. “Continuation methods for approximate large scale object sequencing”. In: *Machine Learning* 108.4 (Apr. 2019), pp. 595–626. ISSN: 1573-0565.

[31] N. Flammarion, C. Mao, and P. Rigollet. “Optimal rates of statistical seriation”. In: *Bernoulli* 25.1 (2019), pp. 623–653.

[32] F. Fogel, A. d’Aspremont, and M. Vojnovic. “Spectral Ranking Using Seriation”. In: *J. Mach. Learn. Res.* 17.1 (Jan. 2016), pp. 3013–3057. ISSN: 1532-4435.

[33] F. Fogel et al. “Convex Relaxations for Permutation Problems”. In: *SIAM Journal on Matrix Analysis and Applications* 36.4 (2015), pp. 1465–1488.

[34] E. Forsyth and L. Katz. “A Matrix Approach to the Analysis of Sociometric Data: Preliminary Report”. In: *Sociometry* 9.4 (1946), pp. 340–347. ISSN: 00380431.

[35] D. Fortin. “An Optimal Algorithm To Recognize Robinsonian Dissimilarities”. In: *Journal of Classi cation* 31 (Apr. 2014).

[36] D. Fortin. “Clustering Analysis of a Dissimilarity: a Review of Algebraic and Geometric Representation”. In: *Journal of Classi cation* 37 (2020), pp. 180–202.

[37] D. Fortin. “Robinsonian Matrices: Recognition Challenges”. In: *Journal of Classi cation* 34 (2017), pp. 191–222.

[38] A. Frieze and R. Kannan. “Quick Approximation to Matrices and Applications”. In: *Combinatorica* 19 (Feb. 1999), pp. 175–220.

[39] M. Ghandehari and J. Janssen. “An Optimization Parameter for Seriation of Noisy Data”. In: *SIAM Journal on Discrete Mathematics* 33.2 (2019), pp. 712–730.

[40] M. Ghandehari and J. Janssen. “Graph sequences sampled from Robinson graphons”. In: (2020). arXiv: 2005.05253 [math.CO].

[41] G.R. Grimmett and D.R. Stirzaker. *Probability and random processes*. Vol. 80. 391. Oxford university press, 2001.

[42] M. Grotschel, M. Junger, and G. Reinelt. "A Cutting Plane Algorithm for the Linear Ordering Problem". In: *Operations Research* 32.6 (1984), pp. 1195–1220. ISSN: 0030364X, 15265463.

[43] F. Harary. *Proof techniques in graph theory : proceedings of the second Ann Arbor Graph Theory Conference, February 1968*. 1969.

[44] J. A. Hartigan. *Clustering Algorithms*. 1975.

[45] H. Hatami, L.M. Lovász, and B. Szegedy. "Limits of locally–globally convergent graph sequences". In: *Geometric and Functional Analysis* 24 (2014), pp. 269–296.

[46] W. Hoeffding. "The strong law of large numbers for U-statistics". In: Institute of Statistics mimeograph series 302. North Carolina State University. Dept. of Statistics, 1961.

[47] P. Hoff, A. Raftery, and M. Handcock. "Latent space approaches to social network analysis". In: *J. Amer. Statist. Assoc.* 97.460 (2002), pp. 1090–1098. ISSN: 0162-1459.

[48] L.J. Hubert. "Problems of seriation using a subject by item response matrix." In: *Psychological Bulletin* 81 (1974), pp. 976–983.

[49] L.J. Hubert. "Seriation using asymmetric proximity measures". In: *British Journal of Mathematical and Statistical Psychology* 29 (1976), pp. 32–52.

[50] T.J.R. Hughes. *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*. Dover, 2000.

[51] J. Janssen and Z. Zhang. "Uniform Embeddings for Robinson Similarity Matrices". In: *Algorithms and Data Structures*. Ed. by Anna Lubiw, Mohammad Salavatipour, and Meng He. Cham: Springer International Publishing, 2021, pp. 499–512. ISBN: 978-3-030-83508-8.

[52] G. Karpilovsky. *The Schur multiplier / Gregory Karpilovsky*. English. Clarendon Press; Oxford University Press Oxford [Oxfordshire]: New York, 1987, xiv, 302 p. ISBN: 0198535546.

[53] D. Kendall. “A mathematical approach to seriation”. In: *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 269 (1970), pp. 125–134.

[54] S Kulczynski. “Die Pflanzenassoziationen der Pieninen”. In: *Bulletin International de l'Academie Polonaise des Sciences et des Lettres, Classe des Sciences Mathematiques et Naturelles, B (Sciences Naturelles)* II (1927), pp. 57–203.

[55] D. Kunszenti-Kovács, L. Lovász, and B. Szegedy. “Measures on the square as sparse graph limits”. In: *Journal of Combinatorial Theory, Series B* 138 (2019), pp. 1–40. ISSN: 0095-8956.

[56] M. Laurent and M. Seminaroti. “A Lex-BFS-based recognition algorithm for Robinsonian matrices”. In: *Discrete Applied Mathematics* 222 (2017), pp. 151–165. ISSN: 0166-218X.

[57] M. Laurent and M. Seminaroti. “Similarity-First Search: A New Algorithm with Application to Robinsonian Matrix Recognition”. In: *SIAM Journal on Discrete Mathematics* 31.3 (2017), pp. 1765–1800.

[58] M. Laurent, M. Seminaroti, and S. Tanigawa. “A Structural Characterization for Certifying Robinsonian Matrices”. In: *Electronic Journal of Combinatorics* 24 (Jan. 2017).

[59] M. Laurent and S. Tanigawa. “Perfect elimination orderings for symmetric matrices”. In: *Optimization Letters* 14.2 (Mar. 2020), pp. 339–353. ISSN: 1862-4480.

[60] C. Lekkeikerker and J. Boland. “Representation of a finite graph by a set of intervals on the real line”. In: *Fundamenta Mathematicae* 51 (1962), pp. 45–64.

[61] J. K. Lenstra. “Technical Note—Clustering a Data Array and the Traveling-Salesman Problem”. In: *Operations Research* 22.2 (1974), pp. 413–414.

[62] I. Liiv. “Seriation and matrix reordering methods: An historical overview”. In: *Statistical Analysis and Data Mining: The ASA Data Science Journal* 3.2 (2010), pp. 70–91.

[63] J. Lindenstrauss and A. Pelczynski. “Absolutely summing operators in $\mathcal{L}_p$-spaces and their applications”. eng. In: *Studia Mathematica* 29.3 (1968), pp. 275–326.

[64] R.L. Ling. “A Computer Generated Aid for Cluster Analysis”. In: *Commun. ACM* 16.6 (June 1973), pp. 355–361. ISSN: 0001-0782.

[65] P.J. Looges and S. Olariu. “Optimal greedy algorithms for indifference graphs”. In: *Computers & Mathematics with Applications* 25.7 (1993), pp. 15–25. ISSN: 0898-1221.

[66] L. Lovász. *Large Networks and Graph Limits.* Vol. 60. Colloquium Publications. American Mathematical Society, 2012, pp. I–XIV, 1–475. ISBN: 978-0-8218-9085-1.

[67] L. Lovász and B. Szegedy. “Limits of dense graph sequences”. In: *Journal of Combinatorial Theory, Series B* 96.6 (2006), pp. 933–957. ISSN: 0095-8956.

[68] L. Lovász and B. Szegedy. “Szemerédi’s lemma for the analyst”. In: *Geom. Funct. Anal.* 17.1 (2007), pp. 252–270. ISSN: 1016-443X.

[69] R. Ma, T.T. Cai, and H. Li. “Optimal Permutation Recovery in Permuted Monotone Matrix Model”. In: *Journal of the American Statistical Association* 116.535 (2021), pp. 1358–1372.

[70] C. Mao, A. Pananjady, and M.J. Wainwright. “Towards optimal estimation of bivariate isotonic matrices with unknown permutations”. In: *The Annals of Statistics* 48.6 (2020), pp. 3183–3205.

[71] W.T. McCormick, P.J. Schweitzer, and T.W. White. “Problem Decomposition and Data Reorganization by a Clustering Technique”. In: *Operations Research* 20.5 (1972), pp. 993–1009.

[72] I. Miklós, I. Somodi, and J. Podani. “Rearrangement of ecological data matrices via Markov chain Monte Carlo simulation”. In: *Ecology* 86.12 (2005), pp. 3398–3410.

[73] B.G. Mirkin, H.L. Beus, and S.N. Rodin. *Graphs and Genes*. Biomathematics. Springer Berlin Heidelberg, 1984. ISBN: 9783540126577.

[74] T. Mishura. "Cut norm discontinuity of triangular truncation of graphons". In: *Linear Algebra and its Applications* 650 (2022), pp. 26–41. ISSN: 0024-3795.

[75] C. Mueller, B. Martin, and A. Lumsdaine. "A comparison of vertex ordering algorithms for large graph visualization". In: *2007 6th International Asia-Paci c Symposium on Visualization*. 2007, pp. 141–148.

[76] G.J. Murphy. "Chapter 2 - C*-Algebras and Hilbert Space Operators". In: *C*{Algebras and Operator Theory*. Ed. by Gerard J. Murphy. San Diego: Academic Press, 1990, pp. 35–76. ISBN: 978-0-08-092496-0.

[77] A. Pananjady, M.J. Wainwright, and T.A. Courtade. "Linear Regression With Shuffled Data: Statistical and Computational Limits of Permutation Recovery". In: *IEEE Transactions on Information Theory* 64.5 (2018), pp. 3286–3300.

[78] W. M. Flinders Petrie. "Sequences in Prehistoric Remains". In: *The Journal of the Anthropological Institute of Great Britain and Ireland* 29.3/4 (1899), pp. 295–301. ISSN: 09595295.

[79] A. Pietsch. "Absolutely p-summing maps in normalized spaces". ger. In: *Studia Mathematica* 28.3 (1967), pp. 333–353.

[80] A. Recanati, T. Brüls, and A. d'Aspremont. "A spectral algorithm for fast de novo layout of uncorrected long nanopore reads". In: *Bioinformatics* 33.20 (June 2017), pp. 3188–3194. ISSN: 1367-4803.

[81] W. S. Robinson. "A Method for Chronologically Ordering Archaeological Deposits". In: *American Antiquity* 16.4 (1951), pp. 293–301.

[82] M. Seminaroti. "Combinatorial algorithms for the seriation problem". English. PhD thesis. Tilburg University, 2016. ISBN: 9789056684969.

[83] N. J. A. Sloane and F. Harary. "Proof Techniques in Graph Theory". In: *Mathematics of Computation* 24 (1970), p. 997.

[84] Peter Henry Andrews Sneath and Robert R. Sokal. *Numerical Taxonomy: The Principles and Practice of Numerical Classi cation*. W. H. Freeman and Co., 1973.

[85] E. Szemerédi. "Regular partitions of graphs". In: *Problemes combinatoires et theorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*. Vol. 260. Colloq. Internat. CNRS. CNRS, Paris, 1978, pp. 399–401.

[86] V.V. Vazirani. "Approximation Algorithms". In: *Approximation Algorithms* (2003).

[87] J.B. Walsh. *Knowing the odds*. Vol. 139. Graduate Studies in Mathematics. An introduction to probability. American Mathematical Society, Providence, RI, 2012, pp. xvi+421. ISBN: 978-0-8218-8532-1.

[88] G. Watson. "Estimating Hadamard operator norms, with application to triangular truncation". In: *Linear Algebra and its Applications* 216 (1995), pp. 97–110.

[89] W. Zhou. "Triangular truncation and its extremal matrices". In: *Numerical Linear Algebra with Applications* 23 (Mar. 2016), pp. 642–655.

Appendix A

STATEMENTS OF USED RESULTS

Theorem A.0.1 (Azuma's Inequality). *Suppose that $\{X_k\}_{k\geq 0}$ is a martingale or supermartingale such that for all k, we have that $|X_k - X_{k-1}| \leq c_k$ almost surely for some positive real number c_k. Then, for all $N \in \mathsf{N}$ and $\epsilon > 0$,*

$$\mathsf{P}(X_N - X_0 \geq \epsilon) \leq \exp\left(\frac{-\epsilon^2}{2\sum_{k=1}^{n} c_k^2}\right).$$

Theorem A.0.2 (Martingale Convergence Theorem). *Suppose that $\{X_k\}_{k\geq 0}$ is a submartingale such that $\sup_k \mathsf{E}(|X_k|) < \infty$. Then there exists a nite random variable X_∞ such that*

$$\lim_{n\to\infty} X_n = X_\infty \text{ a.s.}$$

Lemma A.0.3 (Borel-Cantelli Lemma). *Let $\{E_n\}_{n\geq 1}$ be random events such that*

$$\sum_{n=1}^{\infty} \mathsf{P}(E_n) < \infty.$$

Then, the probability that in nitely of these events occur is 0; equivalently,

$$\mathsf{P}\left(\limsup_{n\to\infty} E_n\right) = 0.$$

Theorem A.0.4 (Perron-Frobenius Theorem). *Let A be an $n\times n$ positive matrix. Then there exists an eigenvector $v = (v_1, \ldots, v_n)$ of A such that $v_i > 0$ for all $1 \leq i \leq n$ and such that its associated eigenvalue $\lambda > 0$ is the largest eigenvalue of A in absolute value.*